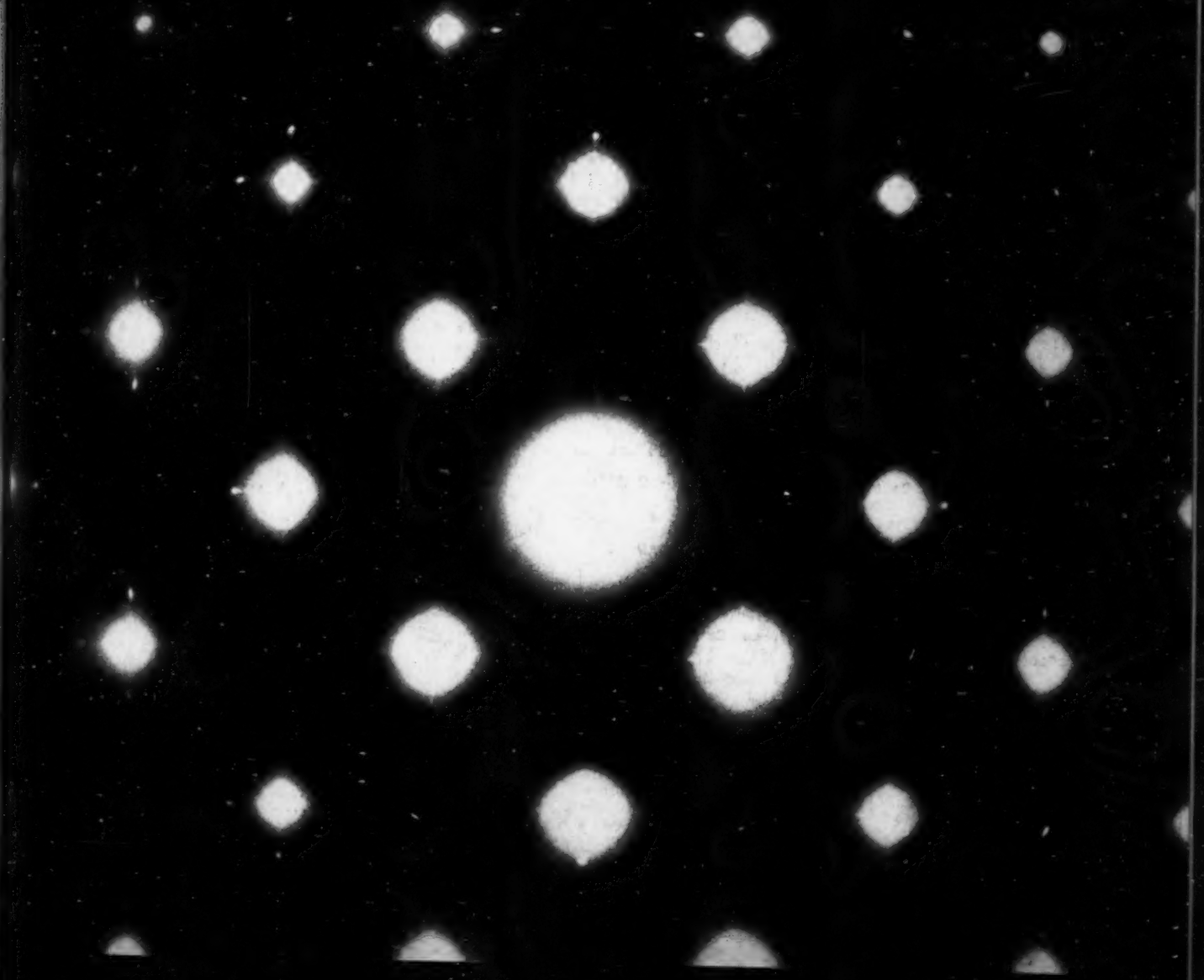
JOURNAL OF APPLIED PHYSICS



JOURNAL OF APPLIED PHYSICS

Vol. 24, No. 2 FEBRUARY, 1953

In This Issue

	Page
Coupling Analog for Nonlinear Systems with More Than One Degree of Freedom — Carl A. Ludeke and Ronald T. Evans	119
Investigation of Back Diffusion of Photoelectrons in Various Standard Gases As It Affects Secondary Electron Emission Coefficients — J. K. Theobald	123
Detection of Weak Signals by Correlation Methods — Philip Rudnick	128
Time-Resolved Spectroscopy of Ultraspeed Pellet Luminosity — William A. Allen and Earle B. Mayfield	131
Direct Determination of the Flow Curves of Non-Newtonian Fluids. II. Shearing Rate in the Concentric Cylinder Viscometer — Irvin M. Krieger and Harold Elrod	134
Nature of the Coefficient of Friction — J. T. Burwell and E. Rabinowicz	136
Alternative Method of Solving Hallén's Integral Equation and Its Application to Antennas Near Resonance — Ronold King	140
Motion of a Sphere Moving Parallel to a Plane Boundary — F. C. Kabal	147
Effect of the Rate of Cooling on the Allotropic Transformation Temperatures of Uranium — Pol Duwez	152
Plastic Electrets — H. H. Wieder and Sol Kaufman	156
Thermoelectric Study of the Electrical Forming of Germanium Rectifiers — Makoto Kikuchi and Tomio Onishi	162
Hall Effect Modulators and "Gyrators" Employing Magnetic Field Independent Orientations in Germanium — W. P. Mason, W. H. Hewitt, and R. F. Wick	166
Screw Dislocations in Thin Rods — J. D. Eshelby	176
Solution of Nonlinear Differential Equations by the Reversion Method — George I. Cohn and Bernard Saltzberg	180
Quasi-Degenerated Modes in High-ε Dielectric Cavities — H. M. Schlicke	187
Flexural Response of a Submerged Solid to a Pulsating Gas Bubble — George Chertock	192
Forced Oscillations with Nonlinear Restoring Force — Chihiro Hayashi	198
Synthesis of Transfer Functions with Poles Restricted to the Negative Real Axis — Louis Weinberg	207
Viscoelastic Properties as Functions of the Distribution of Activation Energies — W. James Lyons	217
Letters to the Editor:	
Shot Noise in Germanium Filaments — R. H. Mattson and A. van der Ziel	222
Shot Noise in Semiconductors — A. van der Ziel	222
Noise in Gas Discharges — A. van der Ziel	223
Pressure-Welded *p-n* Junctions in Germanium — R. G. Shulman and D. M. Van Winkle	224
Magnetic Analog for the Interaction of Dislocation Loops — Edward W. Hart	224
Average Temperatures by Parallel-Connected Thermocouples — Bernard E. Drimmer	225
Hall Effect in Titanium — Georgiana W. Scovil	226
Space Charge Limited Currents between Inclined Plane Electrodes. Approximate Solutions — Henry F. Ivey	227
Evaporated Point Contact Rectifiers — E. G. Roka, C. H. Jackson, and R. P. Ulrich	228
Radiation Damage of Beta-Brass — R. R. Eggleston and F. E. Bowman	229
On the Theory of Prediction of Nonstationary Stochastic Processes — Glenn W. Preston	230
Evidence for the Collapse of Lattice Vacancy Aggregates to Form Dislocation Rings — R. W. Hoffman, F. J. Anders, and E. C. Crittenden, Jr.	231
Theory of Wave-Guide-Fed Slots Radiating into Parallel Plate Regions — L. Lewin	232
On the Zr-Hf System — R. B. Russell	232
New Dispenser Type Thermionic Cathode — Roberto Levi	233
Ball of Fire Discharge with a Pronounced Current Saturation — G. K. Medicus	233
Index to Advertisers	xxviii

Elmer Hutchisson, *Editor*

Earle C. Gregg, *Assistant Editor*

Board of Editors

Ray E. Bolz
Lyle D. Borst
Wallace J. Eckert
Robert D. Heidenreich
Leonard B. Loeb
W. James Lyons
Louis A. Pipes
Roman Smoluchowski
W. R. Smythe

Publication of the

American Institute of Physics

OFFICERS

George R. Harrison, *Chairman*
George B. Pegram, *Treasurer*

GOVERNING BOARD

Karl K. Darrow
G. J. Dienes
S. A. Goudsmit
G. P. Harnwell
George R. Harrison
F. V. Hunt
Rudolph Kingslake
Hugh S. Knowles
C. C. Lauritsen
F. W. Loomis
William F. Meggers
George B. Pegram
I. I. Rabi
Duane Roller
Frederick Seitz
Francis G. Slack
John C. Steinberg
Richard M. Sutton
Mark W. Zemansky

Consultant on Printing

Melvin Loos

ADMINISTRATIVE STAFF

Henry A. Barton, *Director*
Wallace Waterfall, *Executive Secretary*
Theodore Vorburger, *Advertising Manager*
Ruth F. Bryans, *Publication Manager*
Betty H. Goodfriend, *Assistant Secretary*
Eileen B. Neuberger, *Circulation Manager*
Kathryn Setze *Assistant Treasurer*

The Journal of Applied Physics, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as Physics. The Journal of Applied Physics publishes editorials and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio. Unless otherwise stated, submission of a manuscript is a representation that it has been neither copyrighted, published, nor submitted for publication elsewhere. The authors' institutions are requested to pay a **publication charge** of $8 per page which, if honored, entitles them to 100 reprints without further charge. Instructions will be sent with galley proofs.

Subscriptions, renewals, and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Price	U.S. and Canada	Elsewhere
To members of the American Institute of Physics and Affiliated Societies	$10.00	$11.00
To all others	12.00	13.00

Back Number Prices

Yearly back number rate when complete year is available: $14.00. Single copies: $1.25 each.

Changes of address should be sent to the American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics. **Advertising rates** supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879. Accepted for mailing at the special rate of postage provided for in paragraph (d-2), section 34.40, P. L. & R. of 1948, authorized May 2, 1932.

Journal
of
Applied Physics

Volume 24, Number 2 February, 1953

A Coupling Analog for Nonlinear Systems with More Than One Degree of Freedom

CARL A. LUDEKE AND RONALD T. EVANS *
Graduate School of Arts and Sciences, University of Cincinnati, Cincinnati, Ohio
(Received August 5, 1952)

A simple electromechanical analog for expressing the coupling between nonlinear oscillatory systems is constructed and checked. The strength of the coupling can be easily controlled by varying the output of a dc amplifier while the type of coupling is determined by the shape of an aperture in an integrating cylinder. The device has general applicability to analogs in which the motion of a component part can be translated into the motion of a mask before an integrating cylinder with a selected aperature.

INTRODUCTION

IN a previously reported paper,[1] an analog element for solving nonlinear differential equations of the second order for systems of one degree of freedom, was described and utilized. If we wish to tackle nonlinear systems with more than one degree of freedom, consideration must be given to the coupling, linear or nonlinear, between the degrees of freedom. This paper deals with the construction and testing of such a device.

THE INTEGRATING CHAMBER

Since the photointegrating device proved successful in measuring displacements and introducing various nonlinearities in the original analog,[1] this idea was retained in the coupling analog. In order to conserve space and reduce the moment of inertia of the mask support, a cylinder 12.7 cm in diameter, 15.3 cm high, and having an aperture of 2.4–4.6 cm was tried as the integrating chamber. The phototube was inserted from the top of the cylinder 3.2 cm from the wall, with the sensitive surface facing the wall and directly above the aperture. The bottom of the aperture was 1.8 cm from the bottom of the cylinder. The inside of the cylinder was painted a flat white to obtain diffusivity and reflectivity. Experiments were then performed to determine how closely the cylinder approximated the sphere as an integrating device. Various phototubes were used with appropriate light sources as determined by the wavelength sensitivies of the phototubes. The aperture was covered by a straight-edged card which could be moved between the cylinder and the mask. The card was calibrated in units which were one-tenth of the total length of the aperture. The phototube was connected as shown in Fig. 1, and the potential across the load resistance was plotted against the percentage of the aperture exposed. The results of these measurements are shown in Fig. 2. As can be seen from the graphs, various linear combinations of phototubes and light sources can be used with the cylinder for the purposes of integration. A light source was next constructed for the cylinder, and the cylinder was mounted as an integral part of the light source as shown in Fig. 3. This sytem was then mounted above the analog with the shaft and axis of the cylinder in line. A mask was then constructed from balsa wood and attached to the shaft of the analog so that it moved past the aperture of the cylinder. A 100-watt frosted bulb was used for the light source and an RCA 925 phototube for the detecting device.

THE DC AMPLIFIER

It was necessary to amplify the output of the phototube in order to feed enough energy back into the analog to get the desired results. For this purpose, a dc amplifier was designed and constructed. A schematic diagram of the amplifier is shown in Fig. 4. Since the energy fed back into the system is a function of the relative displacement between two shafts, the amplifier had to be able to measure the difference potential of the two phototubes. Drift stability is an important requirement

* Parts of this paper were submitted by Mr. Evans as partial fulfillment of the requirements for the degree of Master of Science, University of Cincinnati, 1951.

[1] C. A. Ludeke and C. L. Morrison, J. Appl. Phys. (to be published).

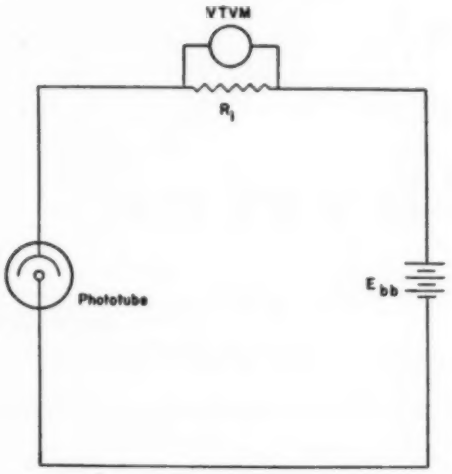


FIG. 1. Integrating cylinder test circuit.

of this type of amplifier so the circuit was designed to give the maximum drift stability for the least cost. Small ripple voltages can be tolerated since the inertia of the analog filters these effects out. Therefore, the filter circuit of the amplifier power supply need not be too extensive. To insure greater stability, the filament circuit of the final stage was grounded through the center tap of the filament winding. This required that the first stage be operated at a negaitve potential with respect to ground. A selenium rectifier with the positive

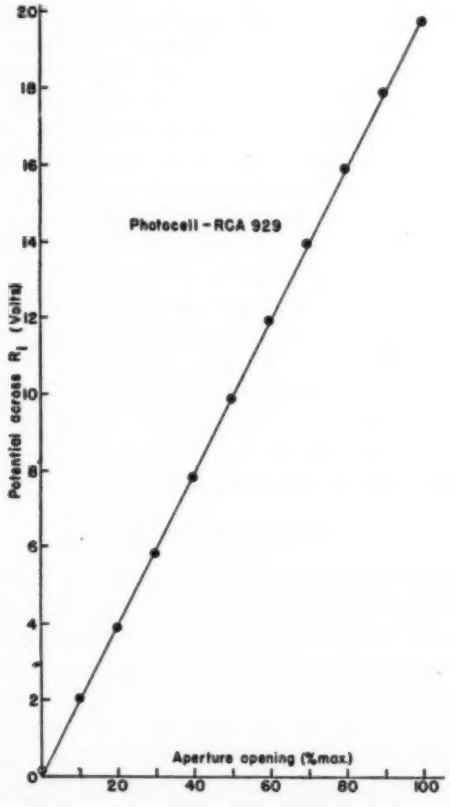


FIG. 2. Linearity of integrating cylinder.

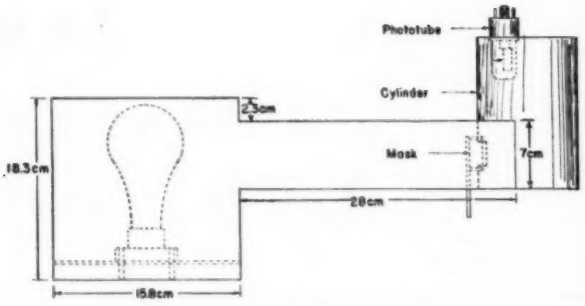


FIG. 3. Integrating cylinder with light source.

terminal connected to ground and the negative terminal connected to the cathode circuit of the first stage was used to obtain the proper operating potentials for this stage. The vacuum tubes selected were the 6SN7-GT for the input and two 6A3's for the output. To meet the high dc requirements of the amplifier, a 5U4-G full-wave vacuum rectifier was selected for the power supply. The power transformer used is a Stancor P-4081. The output stage of the amplifier was designed first, and then the first stage was designed to meet the requirements of the output stage. From the characteristic curves of the 6A3, the quiescent point chosen was 250-volts plate potential and 60-ma plate current. The bias for the 6A3 which will satisfy these conditions is -50 volts. The plate load resistance required to give the proper plate potential was calculated to be 1500 ohms. The first stage was then designed to produce the -50 volts bias required by the 6A3. The grid leak resistance of the 6A3 was selected so that the potential drop across it would maintain the plate of the 6SN7-GT at -50

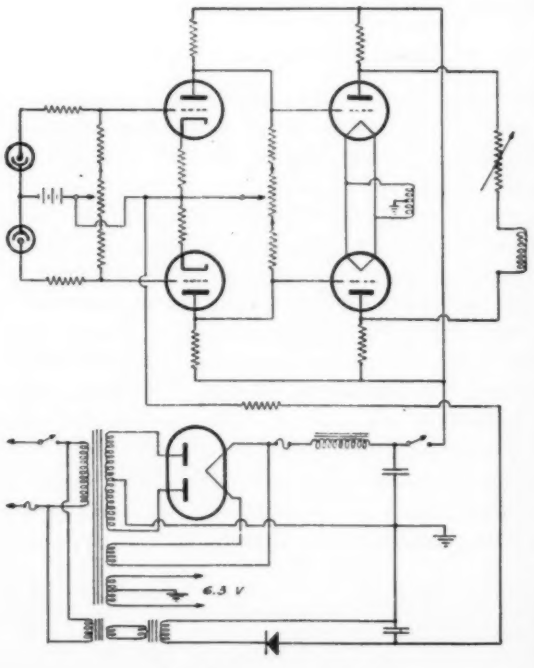


FIG. 4. Diagram of amplifier.

volts when the current in it was 1 ma. The potential of the cathode of the 6SN7-GT is −135 volts so the resistance must be 85 000 ohms. There is a 50 000-ohm variable resistance in this circuit for balancing the grid bias potentials of the 6A3's, so the actual resistance required is 60 000 ohms. The 6SN7-GT quiescent point was picked to be −2 volts bias with 4-ma plate current. The plate load resistance was calculated to be 67 000 ohms. The cathode resistor was calculated to be 500 ohms. The grid resistors of the 6SN7-GT's, which are also the load resistors of the phototubes, are two megohms.

The input to the selenium rectifier is isolated from the ac line through two filament transformers with their low voltage terminals connected together. This combination gives an input of 125 volts. A 40μf condenser across the rectifier and transformer secondary furnishes the filtering for this circuit. A 1000-ohm resistor in series with the selenium rectifier circuit is used to drop the potential to the −135 volts desired. The high voltage power supply uses the 5U4-G rectifier feeding into a single section choke input filter. The choke is 4 henries, followed by a 16μf condenser to ground. A choke input filter was used to provide greater regulation. The completed amplifier was tested for drift stability over a period of 15 minutes and was found to drift 0.4 ma during this period. Figure 5 shows a plot of the drift *vs* time for a 15-minute interval. The linearity of the amplifier was next tested by using the integrating cylinder and light source as the input under conditions approximating those of the actual operating conditions. The results of these measurements are shown in Fig. 6. The analog shaft is confined to move through an angle of 0.5 radian which subtends an arc length of 3.5 cm for a radius of 7 cm. The cylinder aperature is 4.6 cm long, so the operating conditions exclude the end regions of the curves as shown in Fig. 5. The linearity in the operating range is sufficient to anticipate accurate results.

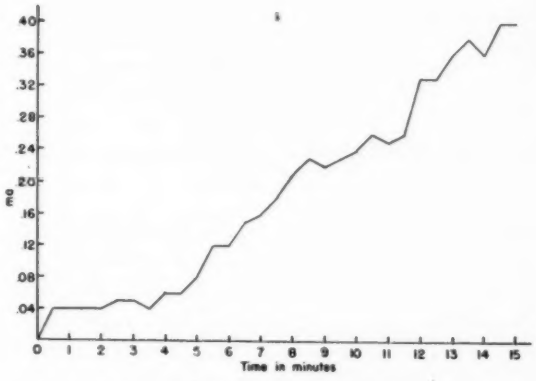


FIG. 5. Drift stability of amplifier.

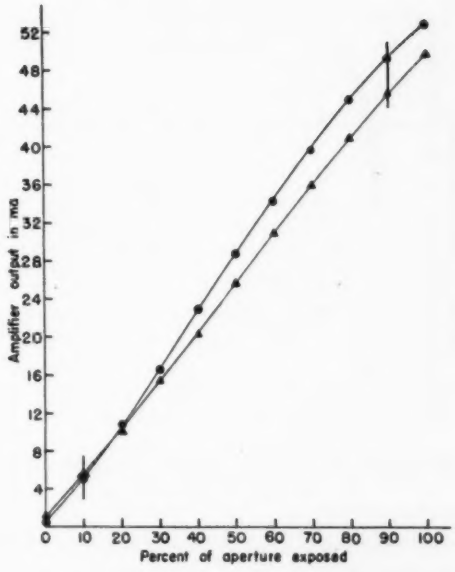


FIG. 6. Linearity of amplifier.

THE COUPLING DEVICE

The output of the amplifier must be used to exert forces on the analog shafts. Since a coil carrying a current and located in a radial magnetic field experiences a torque which is proportional to the current, the output of the amplifier was fed into two such coils, one attached to each of the analog shafts which were to be coupled. Thus if T is the coupling torque between the analog shafts whose displacements are θ_1 and θ_2, we now have a system capable of exerting a torque,

$$T=f_1(\theta_1)-f_2(\theta_2), \tag{1}$$

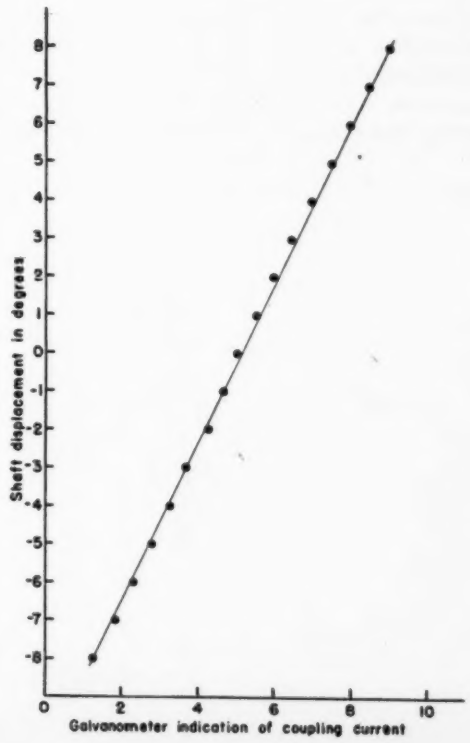


FIG. 7. Linearity of coupling.

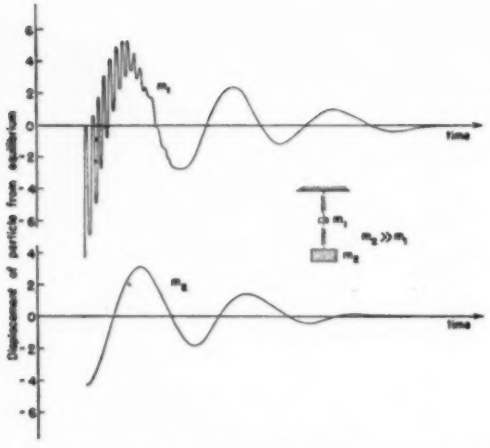


FIG. 8. Solution of a linearly coupled two-degree-of-freedom system. Problem 1.

where the functions f_1 and f_2 are determined by the type of apertures cut in each of the integrating cylinders. To check the system for the case where $f_1(\theta_1)=k_1\theta_1$ and $f_2(\theta_2)=k_2\theta_2$, one shaft was held stationary while the other was given various angular displacements. The results are shown in Fig. 7.

EXAMPLES

Although the device was constructed primarily for the solution of nonlinear problems, it seems appropriate to include for qualitative purposes two examples of linear systems. The first problem and a portion of its solution is shown in Fig. 8. The mass m_2, which is much larger than mass m_1, can be seen to be little affected by the excursions of m_1. In this problem m_1 was displaced and releases before m_2 reached an equilibrium position. The mass m_2 was made large in comparison to m_1 by increasing the moment of inertia of the shaft of the analog representing m_2. The spring constant k_1 was represented by a current of 30 ma in the restoring coil of the analog representing the mass m_1. The spring constant k_2 was the coupling current.

The second example and a portion of the solution is shown in Fig. 9. The two masses m_1 and m_2 are more nearly equal than in the previous example. A further imposed condition is that m_2 as well as m_1 has a constant restoring force applied. Initially m_1 and m_2 were displaced in opposite directions and released.

CONCLUSIONS

It has been shown that coupling terms can be expressed by a simple electromechanical analog. The strength of the coupling can be easily controlled by varying the output of a dc amplifier, while the type of coupling is determined by the shape of the aperture of the integrating cylinder. The device has general applicability to analogs in which the motion of a component part can be translated into the motion of a mask before the aperture of an integrating cylinder.

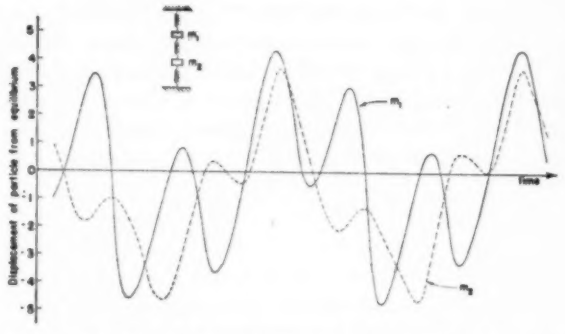


FIG. 9. Solution of a linearly coupled two-degree-of-freedom system. Problem 2.

JOURNAL OF APPLIED PHYSICS VOLUME 24, NUMBER 2 FEBRUARY, 1953

Investigation of Back Diffusion of Photoelectrons in Various Standard Gases as It Affects Secondary Electron Emission Coefficients

J. K. THEOBALD*
Department of Physics, University of California, Berkeley, California
(Received October 15, 1952)

The effect of back diffusion of photoelectrons of two energy distributions in standard molecular and inert gases has been measured, and the results applied to the analysis of the Thomson equation as a semi-empirical formula for correcting the measurements of the secondary emission coefficients in gaseous discharges. The method brings the results of Lauer and Molnar into satisfactory agreement. The effect of electron attachment is briefly discussed.

INTRODUCTION

IN the study of the liberation of secondary electrons from metal surfaces by ions and metastable atoms, the numerical values of the coefficients γi and γm vary widely according to the method of determination. The vacuum techniques employed by Oliphant[1] and Hagstrum[2] give a probability as high as 30 percent for the liberation of a secondary electron by a positive ion or metastable atom of an inert gas incident on a metal surface. The pulsed Townsend discharge at low pressure and high X/p (field strength to pressure ratio) studied by Molnar[3] gave γ of a few percent or less. Lauer[4] studied the pulsed positive wire corona at high pressures, with low X/p at the cathode cylinder, and his values of γ were pressure dependent and less than 10^{-3}.

In the presence of appreciable gas, emitted electrons are returned to the cathode by what is known as back diffusion, and this loss certainly varies with X/p and possibly with p. For example, Molnar operated at an X/p between 70 and 200, and was forced to make independent measurements to correct for back diffusion. At pressures of some tenths of mm of Hg and these X/p, the loss was of the order of 10 percent. In Lauer's apparatus, with p between 25 and 650 mm and X/p at the cathode between 0.5 and 3, the loss may be expected to be much higher. It thus appears logical that the discrepancy can be removed if the probability of the secondary electrons escaping loss by back diffusion is determined. At the low values of X/p found in Lauer's geometry this probability can be directly measured for photoelectrons. This was done for the molecular gases H_2 and N_2 by Bradbury[5] for electrons of one energy distribution.

The present study was undertaken with the following objectives: to determine the back diffusion loss in clean standard gases at various values of X/p and p; to ascertain its dependence on the energy of emission, which varies from less than 1 ev to nearly 10 ev according to the secondary mechanism producing the electrons; to investigate the effect of attachment of electrons to form negative ions on back diffusion in air and oxygen; to check the validity of the Thomson equation as a basis for predicting losses by back diffusion; and, if possible, to obtain data which can be applied in a semi-empirical way to estimate the magnitude of back diffusion loss under different discharge conditions.

EXPERIMENTAL PROCEDURE

The experimental tube, Fig. 1, was a simple diode, with electrodes in the form of nickel disks 3 cm in diameter and 1 cm apart. About 0.5 cm² of the center of the lower electrode was illuminated with ultraviolet light from a quartz mercury arc. These electrodes were outgassed at a red heat produced by electron bombardment for 48 hours, after the tube had been baked at 450°C. The electrodes were surrounded by a nickel screen which shielded the glass walls and, when connected to the anode, acted as a collecting electrode for retarding potential measurements.

Photocurrents were measured in vacuum, in gas at several pressures at X/p ranging from 0.1 up to the onset of ionization by collision, and in vacuum again.

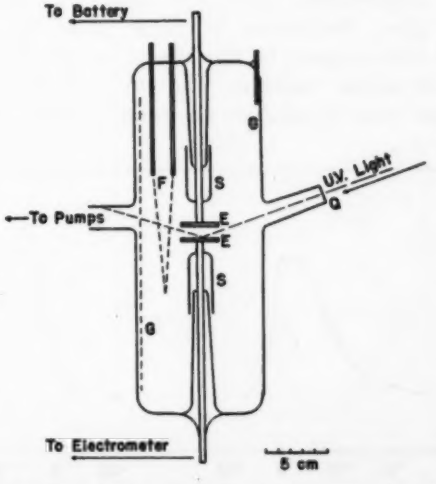


FIG. 1. Experimental tube. *E*, electrodes. *S*, sputtering shields. *F*, outgassing filament. *G*, grounded screen shielding apparatus.

* This project was supported by funds from the U. S. Office of Naval Research.
[1] M. L. Oliphant, Proc. Roy. Soc. (London) A124, 228 (1929); A127, 373 (1930).
[2] H. D. Hagstrum, paper presented at the meeting of the American Physical Society at Berkeley, December, 1951.
[3] J. P. Molnar, Phys. Rev. 83, 933 (1951).
[4] E. J. Lauer, J. Appl. Phys. 23, 300 (1952).
[5] N. E. Bradbury, Phys. Rev. 40, 980 (1932).

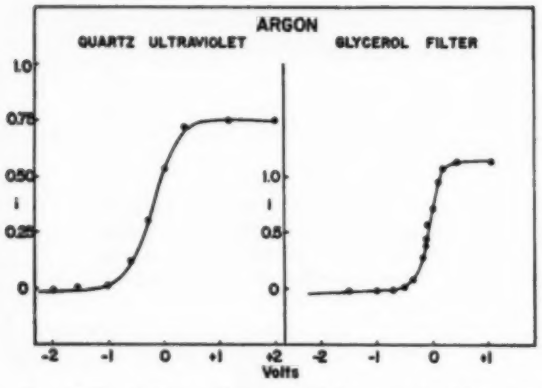


FIG. 2. Typical retarding potential curves, in this case for argon with currents in arbitary units.

If the vacuum current was the same before and after a series of gas measurements it was assumed that the total electron emission had been constant throughout the run.

The photocurrents were kept below 10^{-11} amp to prevent space charge distortion of the field. They were measured by means of a quadrant electrometer and Victoreen resistors, plus a potentiometer which measured the IR drop across the resistor by the null method. The resistors were placed in the cathode circuit to avoid complications introduced by geometrical losses in the anode current at low field strengths, but it was found that the vacuum photocurrent depended on the applied voltage so that saturation was unattainable. This effect appeared to be of the same nature as the thermionic Schottky effect, and amounted to a 10 percent increase in current between 100 and 1000 v. The vacuum photocurrent appropriate to the field was taken to represent the total emission in the presence of gas at that field.

The cathode required no treatment after outgassing to give uniform results in argon and nitrogen. Hydrogen, air, and oxygen produced changes in sensitivity which slowly disappeared when the tube was evacuated. Heavy glow discharges in these gases "*fixed*" the surface with respect to the effects of adsorbed layers and gave stable emission. The effect of oxygen bombardment was to produce an initially high sensitivity which decreased to a very low value in about a day. Retarding potential measurements indicated a fictitiously high electron energy, while rough checks with filters indicated a threshold of 2300A. Hydrogen bombardment produced a satisfactory cathode surface.

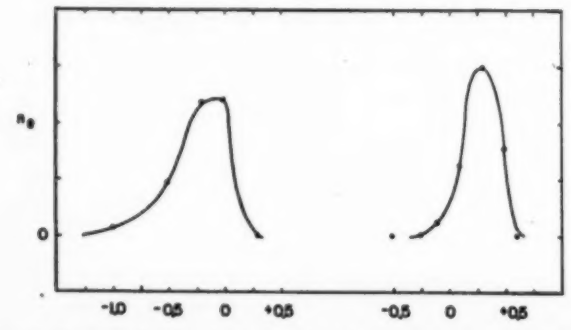


FIG. 3. Energy distribution curves of photoelectrons derived from the data of Fig. 2 with energy increasing toward the left.

Two energy distributions of photoelectrons were obtained by filtering the light through glycerol to eliminate wavelength shorter than 2300A, giving a mean energy of about 0.2 ev, and by using the unfiltered light, which gave an asymmetric distribution whose mean energy is estimated as 0.6 ev. Other liquids were tried in the filter cell, but all seemed to give uneven transmission over the hours required for a run. Attempts to utilize the low work functions of the alkali metals were unsuccessful. A film of potassium gave good stability and retarding potential data, but currents in argon from this surface had an almost linear current-voltage characteristic of a form and magnitude to be expected from negative ions rather than electrons and were considered unrepresentative.

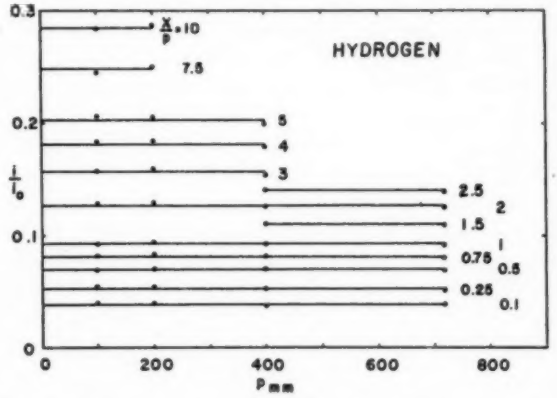


FIG. 4. Characteristic curves for the ratio of the measured current to the total emitted current in vacuum plotted against pressure, in this case for hydrogen at various field strength to pressure ratios. Note the flat characteristics of these ratios, indicating sensible independence of gas pressure.

The argon used was obtained from Airco as "*spectroscopically pure*" in two-liter glass flasks, and gave good results when freshly opened. Commercial hydrogen and nitrogen were passed over hot copper, dried in a liquid nitrogen trap, and passed over incandescent tungsten filaments to remove residual oxygen. Air and tank oxygen were only dried, since few impurities can compete with oxygen for the purpose of this work.

THE THOMSON EQUATION[6]

The Thomson equation is the only mathematical approach to the problem of back diffusion which involves no assumptions as to elastic impacts between electrons and atoms or molecules. Inelastic impacts cannot be neglected, since some secondary electrons have energy considerably in excess of the random

[6] J. J. Thomson, *Conduction of Electricity Through Gases*, third edition.

energy in equilibrium with X/p and must lose energy after entering the gas, whereas the assumption of elastic impacts implies a steady gain of energy as the gap is traversed.

The equation is derived as follows: the number of electrons striking the cathode per cm² per second is given by $\frac{1}{4}nC$, where n is the number density of electrons near the cathode and C is their average velocity. The current density in the gas is $j=nev$, and the total emission corresponds to a current j_0. Then $j-j_0=\frac{1}{4}nCe$, and by substitution, $j/j_0=4v/(C+4v)$.

The approximate nature of this expression is evident. C must be determined from the mean energy of the emitted electrons as shown by Loeb and Bradbury,[5] and this determination involves a relation between rms and average velocity which depends on the form of the energy distribution. Since the electrons are not in equilibrium with the field near the cathode in most cases, both v and n are approximate, n being taken from the measured j and the drift velocity v (as determined by Nielsen for example) for electrons in equilibrium with X/p.

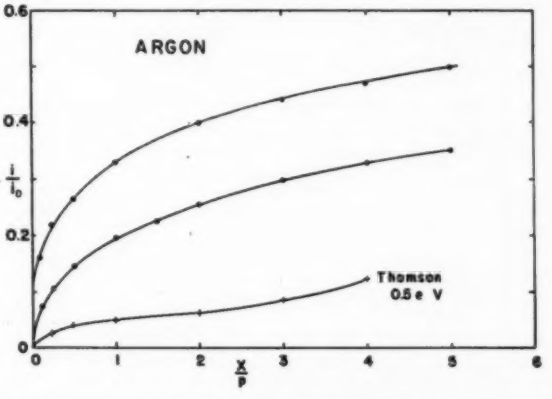


FIG. 5. Current ratios in argon plotted against the ratio X/p. The upper curve is for lower energy electrons, the middle curve for the higher energy electrons represented in Fig. 3. The lowest curve is that computed from Thomson's theory, using an average emitted energy 0.5 v and Nielsen's drift velocities.

RESULTS AND DISCUSSION

The retarding potential curves and the energy distributions derived therefrom, Figs. 2 and 3, are characteristic of the two energy groups in the gases studied, with the exception of air and oxygen, for which these data could not be obtained. The change in work function resulting from hydrogen bombardment is too small to be resolved by these rather crude measurements. The surface was highly sensitized by this bombardment, but the effect seems to have been an increase in the effective area of emission and/or transmission coefficient rather than a reduction of work function. The reproducibility of the vacuum current was usually within a few percent, so the slopes of the retarding potential curves can give only an approximate energy distribution. The error in

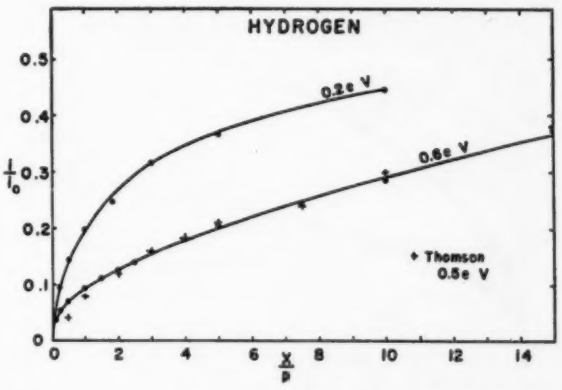


FIG. 6. Current ratio curves plotted against X/p for hydrogen for the two photoelectric energy distributions in hydrogen. Note that the Thomson theory assuming 0.5-v electrons falls closely on the observed curve for 0.6-v electrons except at low X/p.

contact potential determination from these curves is about 0.1 v. Contact potential varied with cathode history, and the variations were not correlated with changes in emission, since in general these depend on the gross average for a surface while emission depends on a few very active centers.

The currents in gas were compared to the vacuum currents at the same field strength, and the ratio i/i_0 of these currents is shown plotted against pressure for hydrogen in Fig. 4, which is typical for the other gases as well. *This ratio is seen to depend only on* X/p and not on p. The pressures studied were sufficiently high that the electron mean free path is very small compared to the electrode separation. From these plots, average values of i/i_0 were taken, and these are shown plotted against X/p for the two energy distributions in Figs. 5 through 9 for the different gases. Points computed from the Thomson equation using Nielsen's[7] v are shown for comparison. The Thomson equation agrees with the observations for the molecular gases for which

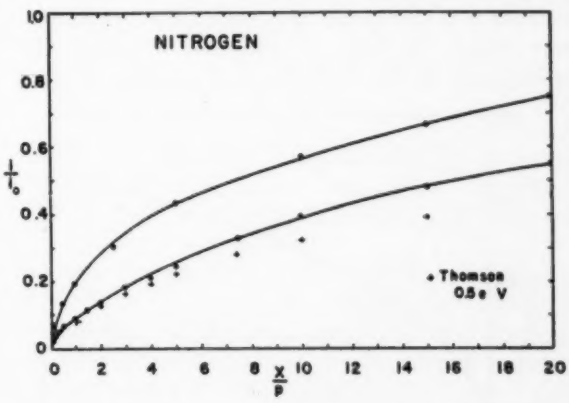


FIG. 7. Current ratio curves plotted against X/p for nitrogen with the upper curves for the photoelectrons of lower energy and the lower curve for those of higher energy. Note that Thomson's equation using 0.5-v photoelectrons coincides at lower X/p but falls below at higher values.

[7] R. A. Nielsen, Phys. Rev. **50**, 950 (1936).

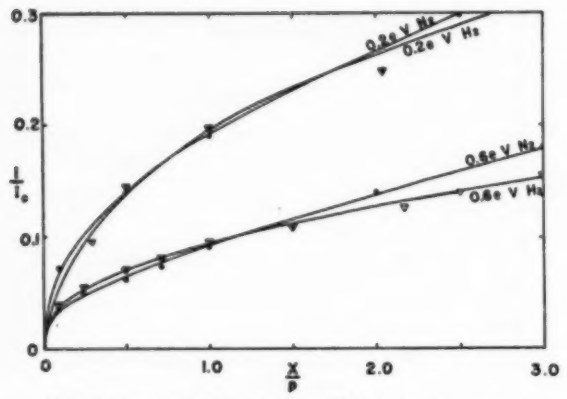


FIG. 8. Current ratios against X/p in hydrogen and nitrogen on an expanded scale to show agreement.

energies were known, but is in error in argon, in which the random energies of electrons in equilibrium with X/p are high, as shown in Table I. The photoelectrons are emitted with energies not far from equilibrium values in the molecular gases and have approximately the equilibrium mobility. Thus data on v give the correct order of magnitude of n near the cathode. In argon, on the other hand, the emitted electrons have much lower energies than the equilibrium values at the X/p used. Consequently, they have a higher drift velocity v than that assumed from the X/p. They gain energy from the field but not rapidly enough to alter v and n near the cathode. Thus i/i_0 is higher than predicted by the Thomson equation as applied, for back diffusion loss is small at low energy, high v, and lower n. The Thomson equation should, however, fit fairly well for secondary electrons produced by inert gas ions or metastables, since these electrons have mean energies of several volts. Then the drift velocity values corresponding to X/p are low, n is high, and loss is great but correctly estimated. The energy dependence of back diffusion also indicates in general that if two processes of equal efficiency are acting to produce secondary electrons in a discharge, the mechanism producing the lower energy electrons will play the dominant role in the economy of the discharge. For this insures higher v and lower n leading to relatively less diffusion loss. The curves for argon show that the electrons are not in equilibrium in the zone of back diffusion, for the upcurving in the Thomson curve which is produced by the increase in v as a result of inelastic collisions at X/p greater than 2 is absent from the observed curves. There is also greater agreement between the observed curves for nitrogen and hydrogen than is to be expected from the similarity of the mobilities in those gases, showing again the importance of the emissive energy in back diffusion.

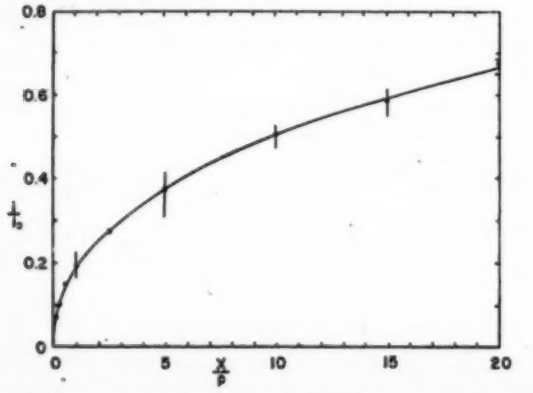


FIG. 9. Current ratio curves against X/p in oxygen indicated by circles. The vertical lines represent the spread of observations obtained in air.

From the viewpoint of the derivation of a proper theory to allow of correction for back diffusion loss, these results indicate what is needed. Young and Bradbury[8] proposed an equation in which the loss is calculated assuming only reflection of electrons in their first encounter with gas atoms or molecules. Correct solution requires a step by step analysis of the scattering of electrons emitted within a given energy distribution over a series of impacts in the gas in which energy is lost by elastic and inelastic impacts with molecules but in turn is gained from the field. The problem doubtlessly poses formidable mathematical-physical difficulties since it falls in the zone where neither straight diffusion or single scattering suffice. It is, however, in essence a diffusion problem.

TABLE I. Bailey's values of electron random energy E in ev for electrons in equilibrium with X/p, as a function of X/p.[a]

Argon		Hydrogen		Nitrogen	Oxygen
X/p	E	X/p	E	E	E
0.125	2.5	0.25	0.08	0.19	0.14
0.195	3	0.5	0.13	0.32	0.24
0.355	4	1	0.23	0.54	0.48
0.525	5	2	0.38	0.76	0.80
0.71	6	5	0.65	1.03	1.08
0.95	7	10	1.1	1.21	1.31
1.25	8	20	1.95	1.5	1.53

[a] Taken from Healey and Reed, *Behavior of Slow Electrons in Gases.*

We conclude that the Thomson equation with Nielsen's v will give results of the right order of magnitude in all cases, with better accuracy if the emissive energy of the secondary electrons is reasonably close to the equilibrium energy at the value of X/p effective in the zone of back diffusion. The departure from equilibrium of the observed data for argon is as extreme as can be expected, and the disagreement between the observations and the Thomson equation amounts to a factor of less than 5. Fast secondary electrons in the molecular gases must lose energy rapidly because of the inelasticity of collisions in those gases, and it is likely that the initial mean energy will be appreciably lowered before the electrons leave the zone of back diffusion.

Toward the end of furnishing approximate data for correction of results for back diffusion loss, Figs. 10 and 11 have been prepared. They show the variation of i/i_0 with the reciprocal of the emissive energy as estimated from the observed data at 0.2 and 0.6 ev and the Thom-

[8] Young and Bradbury, Phys. Rev. 43, 34 (1933).

son equation at higher energies, for argon and hydrogen. By choosing the emissive energy cogent to a particular gas and situation the correction for the X/p in question the values of i/i_0 may be read from the curves. These curves should fit within a factor of two or so for the inert gases and the molecular gases, respectively, for secondary electrons of known initial energy. Applying these corrections to the values of γ obtained by Lauer has the results shown in Table II. The ion-produced secondary electrons are assumed to have a mean energy of 10 ev and the photoelectrons are given 3 ev, estimated from the observations of Hagstrum[8] and Kenty.[9] The corrected γ for argon fall into two groups according to pressure. Lauer reports a small pulse presumably the result of A^+ at the two higher pressures, which could not be resolved from the pulse which resulted from A_2^+ at the lower values of pressure. Conceivably there are two mechanisms active here yielding values of γ in the ratio of 10. The atomic ion is doubtlessly present in greater amount at lower pressures. It should give a γ greater than that for the molecular ion. In He the two ions have γ differing by a factor of less than two and not of 10 as appears to be the case here. However, nothing is known about the relative efficiencies of the two A ions. The corrected coefficients agree reasonably well with those obtained for Mo and Ta by Molnar.

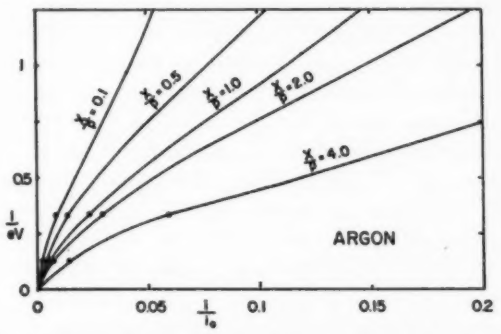


FIG. 10. Current ratio curves plotted against the reciprocal of emissive energy in electron volts for argon at various X/p. These curves permit rough correction of experimental data for back diffusion loss for argon within a factor of 2 and probably apply to the other inert gases over the range given.

In the case of hydrogen the pressure variation of γp persists despite the back diffusion correction. The corrected values again do not disagree greatly with those obtained by Molnar in A at low pressure, where the active photons are probably those of the 1061A resonance line, corresponding to the extensive hydrogen system in that spectral region. The linear pressure variation of the corrected γ may stem from the change of the relative intensities of the component frequencies with the rather wide variation of X/p near the wire, since the photoelectric coefficient is defined as the number of photoelectrons liberated per positive ion created in the discharge.

[9] C. Kenty, Phys. Rev. 44, 891 (1933).

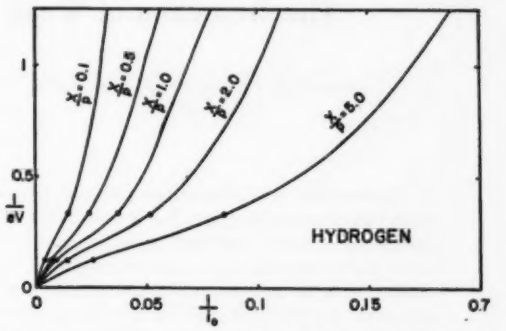


FIG. 11. Current ratio curves for hydrogen plotted against the reciprocal of emissive energy in electron volts at various values of X/p. These curves permit the correction of data for back diffusion for molecular gases with relatively inelastic impacts over the range given.

Loeb and Kunkel have analyzed the effect of electron attachment on back diffusion, with the following results: electron attachment is relatively slow so that it does not alter n at the cathode appreciably. Hence the only effect of attachment is to create a space charge of slow moving negative ions, which lowers the effective field at the cathode. This is not observable at the current densities employed in this work. The expression for the effective field at the cathode is

$$X_e = \frac{9k_iV^2}{32\pi e n_e k_e d^3(1-\bar{x}/d)^3},$$

where k_i = ion mobility, k_e = electron mobility, $\bar{x}$ = average distance traveled by electron before attaching, d = electrode spacing, n_e = electron density, V = applied voltage, and e = electronic charge.

For air at 760 mm, $d=1$ cm, $\bar{x}=0.1$ cm, $n_e=10^5/\text{cm}^3$, and $V=100$ v, $X_e=66$ v/cm. This corresponds to a current density of 3×10^{-9} amp/cm². The oxygen treated cathode used in this work gave a current density of less than 10^{-12} amp/cm².

TABLE II. Lauer's observed γ corrected for back diffusion.

Argon p (mm)	X/p[a]	γ_i	γ_e
25	3.12	7.8×10^{-4}	0.035
50	1.92	7.0×10^{-4}	0.051
100	1.18	4.0×10^{-4}	0.040
200	0.79	0.6×10^{-4}	0.006
400	0.52	0.2×10^{-4}	0.002
Hydrogen		γ_p	
100	1.43	6.9×10^{-4}	0.014
200	1.04	1.5×10^{-4}	0.004
400	0.79	0.74×10^{-4}	0.002
650	0.66	0.49×10^{-4}	0.0015

[a] X/p is the value at the cathode.

ACKNOWLEDGMENT

The writer wishes to express his appreciation of the support and encouragement given by Professor L. B. Loeb, who proposed this problem and under whose direction it was carried out.

The Detection of Weak Signals by Correlation Methods*

PHILIP RUDNICK
University of California, Marine Physical Laboratory of the Scripps Institution of Oceanography, La Jolla, California

(Received July 3, 1952)

A procedure involving autocorrelation is described which, for long averaging time and incoming signals with continuous spectrum, has a detection threshold equal to that of a corresponding conventional system consisting of band pass filter, rectifier, and low pass filter. The autocorrelation procedure is quite complex; a full autocorrelation function is computed and then subjected to a "filtering" operation. A simpler procedure, such as computing autocorrelation coefficients for only one or a few delay times, will in general be less effective. A third process, involving multiplication by a local sinusoidal signal, is also discussed and shown to be closely parallel to the conventional system.

A REPORT with the above title has been issued and given limited distribution.[1] To make the main conclusions more generally available, the following condensation is given here.

The general problem considered is that of recognizing the presence of a small steady signal within a fixed time T after it has been added to a steady background noise in any communication channel. The familiar way of doing this is with a filter and rectifier (see discussion of Fig. 1, below). Autocorrelation and cross correlation have recently received a good deal of attention as alternative procedures.[2,3] The sensitivities of correlation processes have been studied analytically.[4–7] The particular schemes discussed here are chosen to show close parallelism with the familiar filter-rectifier method. The detection thresholds found for these analogous methods are generally alike or comparable, as may be expected.

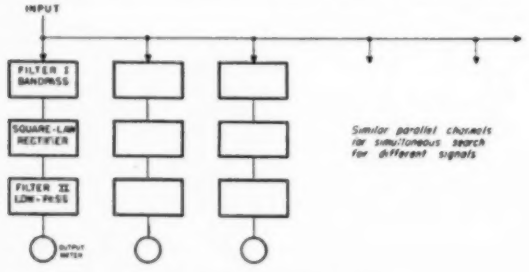


FIG. 1. Filter-rectifier system.

Figures 1, 2, and 3 are block diagrams of three detection processes. The output of each is the continuous indication of a meter whose mean or dc reading changes with presence or absence of the signal. The meter must of necessity also show fluctuations due to background which establish a lower limit to the magnitude of signal which can be detected. To make this threshold determinate it is essential to specify some time interval T within which the signal must be detected. This sets an upper limit on the transient response time of the detection system, and a lower limit on the background frequencies which can be eliminated in the output.

Figure 1 represents a conventional process. Filter I introduces a favorable spectral selectivity. The rectifier generates a dc component which measures input power and can indicate presence or absence of signal. Filter II is low pass, transmitting the dc component while rejecting as much fluctuation as possible, consistent with the requirement or transient time of the system. If the spectral character of the signal is not known in advance, a set of parallel channels may in principle be added, with filters I differing in center frequency, band width, etc.

The system shown in Fig. 2 is, in brief, a set of correlators which jointly compute the full autocorrelation function of the input. These are followed by one or more analyzing networks which respond to components of particular form in this autocorrelation function. Each correlator channel receives two versions of the input differing by a relative time delay τ which is fixed for any one channel but varies from one to another channel. A multiplier forms the product of the displaced inputs which is then averaged by a low pass filter II. The set of dc outputs of the various channels presents the mean displaced product, or by definition the autocorrelation, of the input as a function of τ. Each analyzing network forms an instantaneous linear combination of all these outputs, the set of coefficients in

* This work represents one of the results of research carried out under contract with the Bureau of Ships, Navy Department. Contribution No. 611 from the Scripps Institution of Oceanography.

[1] Philip Rudnick, "The detection of weak signals by correlation methods," Marine Physical Laboratory of the Scripps Institution of Oceanography, La Jolla, California, SIO Ref. 52-26, May 10, 1952. Additional copies, as long as available, may be had upon request.

[2] Lee, Cheatham, and Wiesner, Proc. Inst. Radio Engrs. 38, 1165 (1950).

[3] F. W. Hunt, Proc. Inst. Radio Engrs. 39, 840 (1951).

[4] Carl Eckart, "The measurement and detection of steady A.C. and D.C. signals in noise," University of California, Marine Physical Laboratory of the Scripps Institution of Oceanography, SIO Ref. 51-39, La Jolla, California, October 4, 1951.

[5] R. M. Fano, "Signal-to-noise ratio in correlation detectors," Tech. Rept. No. 186, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, February 19, 1951.

[6] W. B. Davenport, Jr., "Correlator errors due to finite observation intervals," Tech. Rept. No. 191, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, March 8, 1951.

[7] Davenport, Johnson, and Middleton, J. Appl. Phys. 23, 377 (1952).

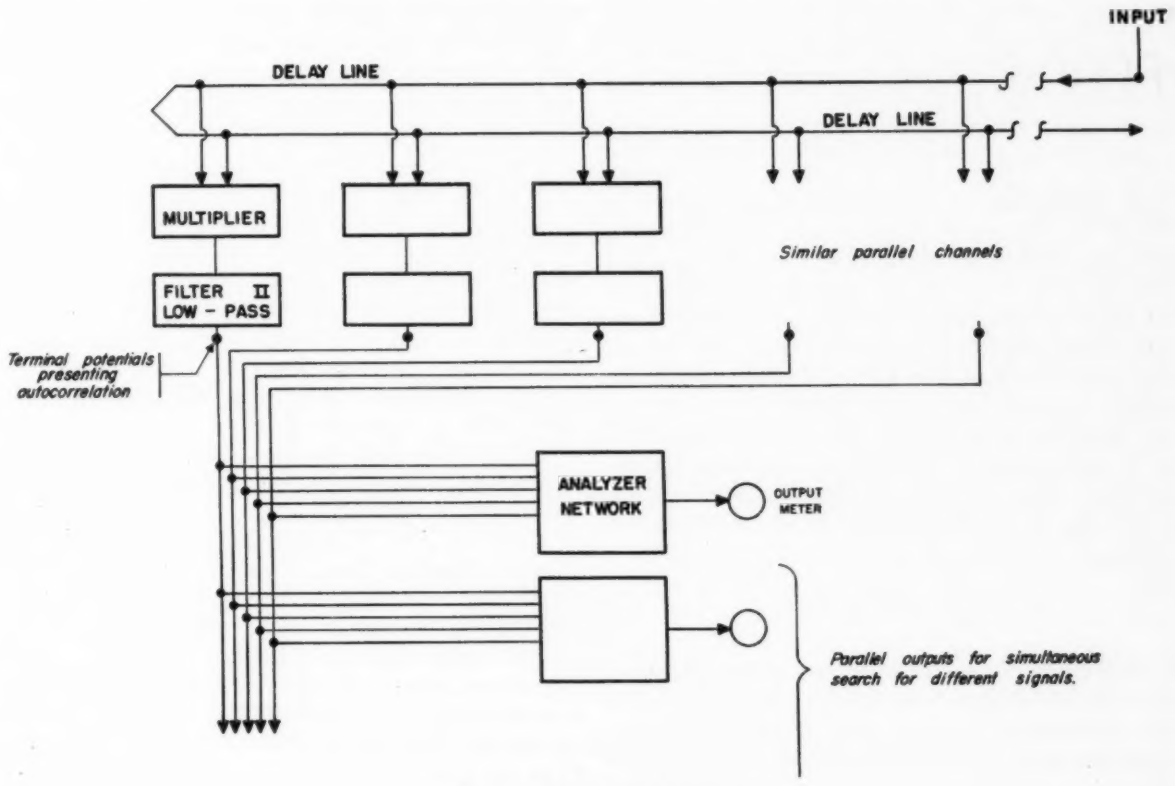


FIG. 2. Autocorrelation system.

this combination being determined by the particular form of component which is to be detected in the correlation function. Again, each filter II must pass, as well as the desired mean or dc output, some unwanted fluctuations which carry through to the final indicating meter.

The filter-rectifier and autocorrelation procedures are quite closely analogous. In each, a quadratic process generates a mean output, which is then isolated as fully as possible by a low pass filter. Each also involves spectral discrimination, in one case preceding and in the other following the quadratic process. For any specific choice of filter I in Fig. 1, a corresponding set of coefficients can be determined for the analyzing network in Fig. 2. This correspondence rests, of course, on the one-to-one relation between power spectra and correlation functions, each being a cosine transform of the other.

Let us consider two systems in which this correspondence between filter I and the analyzing network has been established, and which are also alike in that each quadratic element gives an output of 1 volt per (volt)^2 input and the low pass filters II are identical. We suppose both signal and background entering each system to have continuous spectra and Gaussian statistical properties, and that the signal, when present, is statistically independent of the background. These two systems will then produce identical increments of mean (dc) output when any given signal is added to background at each input. Calculations[1] yield the following results:

increment in mean output due to signal for either system

$$= \int_{-\infty}^{\infty} s(\tau) f_1(\tau) d\tau; \qquad (1)$$

mean squared fluctuation in output for filter-rectifier system

$$= 2 \int\int\int a(\tau-\tau_1) a(\tau-\tau_2) f_1(\tau_1) \times f_1(\tau_2) f_2(\tau) d\tau_1 d\tau_2 d\tau; \qquad (2)$$

mean squared fluctuation in output for autocorrelation

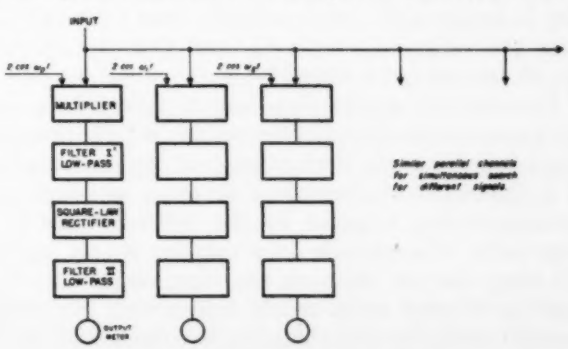


FIG. 3. Low frequency heterodyne system.

system

$$=2\int\int\int a\left(\tau+\frac{\tau_2-\tau_1}{2}\right)a\left(\tau-\frac{\tau_2-\tau_1}{2}\right)$$

$$\times f_1(\tau_1)f_1(\tau_2)f_2(\tau)d\tau_1 d\tau_2 d\tau. \quad (3)$$

In these expressions $s(\tau)$ and $a(\tau)$ are autocorrelation functions of the input signal, and of background, respectively. (The observation time T will be supposed sufficiently long to make the least detectable signal power spectrum density well below background. Then the important fluctuations are those due only to the background.) The functions $f_1(\tau)$ and $f_2(\tau)$ characterize the filters I and II, respectively. Each may be defined as the autocorrelation of the filter output when the input is white noise of unit power spectrum density (normalized over $-\infty \leq f \leq \infty$). An equivalent definition involving the filter response $I(t)$ to unit impulse at $t=0$ is

$$f(\tau)=\int_{-\infty}^{\infty} I(t)I(t+\tau)dt. \quad (4)$$

The right side of Eq. (1) when applied to the filter-rectifier process is a measure of the power (from the signal) transmitted through filter I. Applied to the autocorrelation process, the integral is replaced by a discrete sum which constitutes a definition of the output of the analyzing network in response to a set of inputs represented by $s(\tau)$. The coefficients $f_1(\tau)$ are characteristic of the analyzing network and indicate the way in which it is related to filter I.

The chief result of this discussion is that the expressions given in (2) and (3), although in general differing (usually, if not always, in favor of the filter-rectifier system), become identical in the limit of very low cut-off frequency for filter II, corresponding to very long observation time T. In this case, the common result can be transformed to a very simple expression: mean squared fluctuation for long averaging time, for either system,

$$=4\bar{f}_2\int_{-\infty}^{\infty} A^2(f)\,\mathfrak{f}_1^2(f)df. \quad (5)$$

Here $A(f)\,\mathfrak{f}_1(f)$ represents the power spectrum density, due to background, transmitted by filter I (normalized over the range $-\infty \leqq f \leqq \infty$), and $\bar{f}_2$ is an effective cut-off frequency for filter II.

Thus for long observation time T, we conclude that the conventional filter-rectifier system is fully as sensitive a detector as the alternative involving computation of a complete autocorrelation function, although the instrumentation required by the latter scheme is a large order of magnitude more complex if only signals of a single spectral character are being sought. The two methods become more nearly comparable in instrumental complexity if the search is to cover a wide array of different signals. However, in any case, it seems fair to say that the scheme of Fig. 2 will not be put to practical use without substantial advance beyond the present state of the art of correlator design. The more immediately useful result is the conclusion that simpler autocorrelation procedures, involving only one or a few values of the time delay τ, cannot be expected to do as well as the standard filter-rectifier process.

It may be noted that the comparison of these two systems bears closely on the question posed by Davenport, Johnson, and Middleton,[7] of the comparative time required for the determination of a power spectrum directly by spectral analysis, as against indirectly through the correlation function. These writers do not clearly distinguish between the time required for sufficient data to enter the system, which is an essential time, and the time required to perform successively operations which can in principle be performed simultaneously, or with arbitrarily high speed, which is time required by cost or convenience. The first of these aspects, essential time or quantity of data required, is the only one treated here. If in Fig. 1 the filters I are supposed to be narrow and contiguous, the set of dc output meter readings constitutes a finitely resolved power spectrum of the input, and if Fig. 2 represents a corresponding system receiving the same input, its output meters will give identical dc information. The fluctuations are still given by Eq. (5), or by (2) and (3), and we find the systems equivalent as power spectrum analyzers if the available sample of data is sufficiently large.

Another process, shown in Fig. 3, has also been studied. This is appropriate for a signal whose spectrum is continuous, but confined to a narrow band. The input is multiplied by a local oscillation of single frequency within the signal band, then fed to a low pass filter I′, which acts analogously to filter I in Fig. 1. The succeeding components also correspond. The operation of the multiplier and filter I′ will be called heterodyning to low frequency. It should be remarked, however, that if a dc output were generated by the multiplication, the same operation would be called cross correlation. The calculated over-all performance of Fig. 3 is very closely parallel to that of Fig. 1, except that the least detectable input signal power is greater by a factor $2^{\frac{1}{2}}$ in the low frequency heterodyne process. This is apparently a price that must be paid if the band pass filter I is replaced by a low pass filter I′, which in some circumstances may have a practical advantage.

The superiority of cross correlation with a local sinusoid over simple autocorrelation (with a single delay) for signal detection has been noted in the literature.[2,5] This type of cross correlation has a detection threshold quite comparable to those of the systems of Fig. 1 and Fig. 3.† Simple autocorrelation tends to fall

† For an input proportional to $\cos\omega_0 t$, the threshold input powers are as $1:2:2^{\frac{1}{2}}$ in the order cross correlation: Fig. 1: Fig. 3. In this case Fig. 3 is not inferior to Fig. 1, because the input and local signals agree in phase.

far short of all three; only the more elaborate procedure of Fig. 2 provides a fourth method of comparable sensitivity.

A simple summary of some aspects of these results can be stated in terms of a signal-to-noise improvement ratio. This we define as the quotient of signal-to-noise ratio at output to that at input, when the latter is a power ratio and the former a voltage ratio, of dc increment due to signal, to rms fluctuation due to background. Improvement ratios defined in this way, for the case of narrow-band signal in broad-band noise, with long observation time T, can be expressed[1] (with some simplifying approximations) in terms of three effective band widths: β for input background, $\bar{f}_1$ for filter I (assumed not narrower than the input signal spectrum), and $\bar{f}_2$ for filter II. (In case the action of filter II is replaced by a running arithmetic average using equal weights over a time interval T, then $\bar{f}_2=1/(2T)$. For either filter-rectifier or autocorrelation system (Fig. 1 or 2) the over-all improvement ratio is $\beta/(2\bar{f}_1\bar{f}_2)^{\frac{1}{2}}$. For the filter-rectifier process this may be separated into two factors, $(\beta/2^{\frac{1}{2}}\bar{f}_1)$ and $(\bar{f}_1/\bar{f}_2)^{\frac{1}{2}}$, the first of which can be ascribed to filter I plus rectifier, and the second to filter II. For the autocorrelation process the same over-all improvement may be separated into a factor $(\beta/\bar{f}_2)^{\frac{1}{2}}$ characterizing the effect of a single correlator channel, and a second factor $(\beta/2\bar{f}_1)^{\frac{1}{2}}$ showing the gain afforded by the analyzing network in combining the many correlator outputs.

The over-all improvement ratio $\beta/(2\bar{f}_1\bar{f}_2)^{\frac{1}{2}}$ is increased by reducing the values of $\bar{f}_1$ and $\bar{f}_2$. $\bar{f}_2$ always has a lower limit of order $1/(2T)$ when T is (as before) the allowed detection time. The lower limit of $\bar{f}_1$ will be of the order of either the input signal band width, or $1/(2T)$, whichever is greater. The former case is assumed in the preceding paragraph and leads to over-all improvement proportional to $T^{\frac{1}{2}}$. In the latter case $\bar{f}_2$ is no longer small compared to $\bar{f}_1$, but of the same order, or filter II is omitted altogether, and the over-all improvement is of order $\beta/\bar{f}_1$, proportional to T.

JOURNAL OF APPLIED PHYSICS VOLUME 24, NUMBER 2 FEBRUARY, 1953

Time-Resolved Spectroscopy of Ultraspeed Pellet Luminosity* †

WILLIAM A. ALLEN AND EARLE B. MAYFIELD
Michelson Laboratory, U. S. Naval Ordnance Test Station, Inyokern, China Lake, California
(Received October 21, 1952)

Improved optical instrumentation has resulted in high time-resolution spectrograms of the luminosity associated with ultraspeed aluminum pellets. These data indicate that in many cases the light is emitted in two phases: (1) the ballistic, or air shock, characterized by atomic metal lines and a strong continuum; (2) AlO bands that appear about 150 μsec after the metal is ablated from the pellet.

INTRODUCTION

IN this paper are presented certain experimental data obtained at the U. S. Naval Ordnance Test Station on time-resolved spectroscopy of the incandescence associated with ultraspeed pellets. A time-resolved spectrogram has been published[1] of the wake of an aluminum pellet traveling through air at 5 km/sec. Recently, further experiments have been made with a spectrograph of greater time-resolution.

TIME-RESOLVING SPECTROGRAPH

Figure 1 is a photograph of a high speed camera loaned the Naval Ordnance Test Station by the Los Alamos Scientific Laboratory. This camera uses a 2-ft length of 16-mm film on the inside of a rotating drum mounted on the shaft of an 18 000-rpm Dumore motor. A 45-mm $f/2$ Ektar lens is mounted above the drum with its axis normal to the plane of rotation. The image is reflected 90 deg onto the film by means of a small front-surface mirror. Film velocity is obtained by the following two independent methods:

(1) In the first timing method, millisecond pips are impressed on the edge of the moving film by a specially constructed precision pulse generator that activates a 1/25-w neon bulb mounted inside the film drum. To prevent overlapping timing marks on the film, the timing circuit is gated at zero time for about 7 msec by a commutator switch.

(2) In the second timing method, a small iron bar insert is mounted on the shaft of the motor in an aluminum disk that rotates in a magnetic field produced by current flowing in a pick-up coil. The frequency of the resultant pulses is compared visually on an oscilloscope screen with the frequency from a pre-set audio-oscillator.

The drum camera was converted into a time-resolving spectrograph by a train of two 4-in., 60-deg, DF-2 prisms mounted at minimum deviation. To reduce the

* Work supported by Armament and Mechanics Branches, U. S. Office of Naval Research.
† This paper is based upon a presentation before the December, 1951, Berkeley meeting of the American Physical Society—Earle B. Mayfield, Phys. Rev. 85, 769 (1952).
[1] Allen, Rinehart, White, J. Appl. Phys. 23, 198 (1952).

FIG. 1. High speed drum camera with spectrograph attachment.

effect of slit curvature and to eliminate flare spots, the instrument was provided with a 1-mm focal plane aperture slit and a 1-m field stop in the plane of the object. The spectrograph was focused on a 50-mm line source at a distance of 168 ft. Critical photographic focus was obtained while the drum was rotated at operating speed. The film was exposed at a series of lens settings to a mercury vapor light source. Each lens setting was measured by a micrometer depth guage. Critical focus for a given spectral line was taken as the position that resulted in the narrowest line width as determined by a microphotometer trace of the film. Figure 2 illustrates the variation of the back focal length of the lens with respect to wavelength. Field focusing for a given wavelength was accomplished with the depth guage where its setting was obtained from Fig. 2.

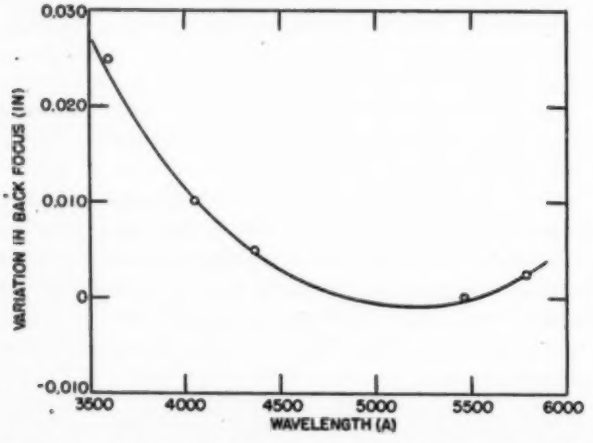


FIG. 2. Variation in focal length of optical system at 168-ft object distance.

Daytime use of a rotating drum camera is predicated upon a fast shutter synchronized with the explosion. For improved firing simultaneity, use was made of electrical detonators designed for seismograph work.‡ As an extra precaution, the detonators were pulsed with several kilovolts although they will fire with only a few volts. A combination fast shutter and high voltage switch, illustrated in Fig. 3, was designed to serve the dual purpose of shuttering the camera and firing the charges. The device consists of a rotating disk powered by a spring. The disk contains a single hole of the necessary aperture. When the trigger solenoid (*A*) is pulsed, the latch (*B*) is released, the disk rotates through 180 deg, and is brought to rest against a brake shoe (*C*). During its rotation, the aperture in the disk aligns at 90 deg with the apertures in the frame of the shutter case. At this angle, one terminal of a high voltage switch has been brought into proximity with its mate

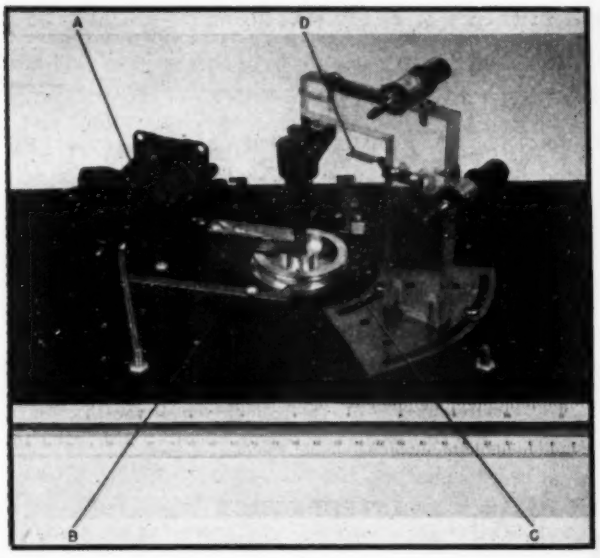


FIG. 3. Combination fast shutter and high voltage switch.

(*D*), and the resultant contact closes the firing circuit. The strength of the spring is chosen to obtain the proper shutter speed. For the 2-in. aperture used, a shutter speed of 0.01 sec is obtained easily.

ULTRASPEED PELLET EXPERIMENT§

The incandescent trail of a 5 km/sec aluminum pellet is used as a line source for the time-resolving spectrograph. Figure 4 is a photograph of a pellet trail across the 1-m field stop. To obtain this photograph, aperture *f*/45 and exposure 1/25 sec, it was necessary to synchronize the still camera shutter with the explosion.

‡ These detonators were manufactured by the Hercules Powder Company under the specifications: Electric Blasting Cap for Seismograph Use, No. 8 Vibrocap, Waterproof, 20 Foot Enamel Wires.

§ Previous work with ultraspeed pellets has been reported in reference 1, p. 132.

Figure 5(a) is a discrete photograph of the aluminum pellet that produced the trail in Fig. 4. This photograph was obtained with a General Radio continuously moving film camera by means of a technique previously reported.[2] A time-resolved spectrogram, Fig. 5(b), was obtained by moving the spectrograph film parallel to the image of the line source, shown in Fig. 4. Since the photographs were obtained by sweeping field images, the horizontal axes of Fig. 5 represent time. The initial portion of the spectrogram is marked by a strong continuum, Al, Mg, and Mn metal lines[3] whose excitation potentials are comparable with the translational energy of the air atoms relative to the pellet. At about 100 μsec, the pellet trail is marked by a null in luminosity where all spectral lines vanish. At about 150 μsec, the AlO bands appear for the first time; the released thermal energy results in a recurrence of the Al I lines at λλ3944 and 3961 shown near the top, or blue end, of the spectrogram. The qualitative light curve obtained by scanning the time-resolved spectrogram with a microphotometer, Fig. 5(c), is not unlike that of an atomic bomb.[4] The luminosity curve is marked by (1) an initial spike, followed by (2) a null, in turn followed by (3) a secondary maximum. The primary maximum is caused by energy interaction between the pellet and the air; the secondary maximum is a result of burning. Pellets composed of a magnesium-lithium alloy[5] do not exhibit the secondary maximum. A spectrogram of the Mg—Li pellet trail is marked by a strong initial continuum and the easily excited lines of Li and Mg. The lines persist for a few μsec while the continuum lasts for several hundred μsec. Faint MgO bands appear at about 200 μsec.

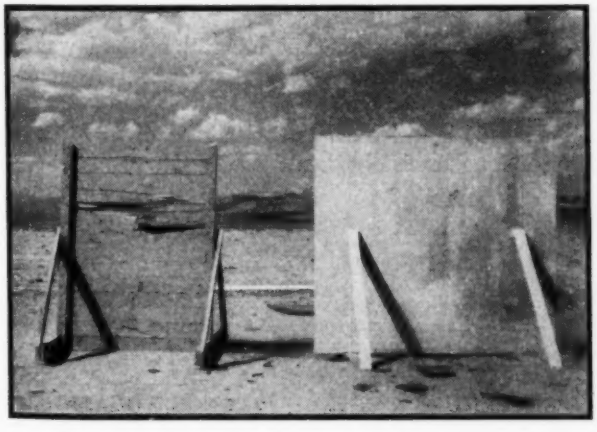


FIG. 4. Ultraspeed pellet traveling past a 1-m field slit. The trail reproduces as a streak resulting from the 5-km/sec velocity.

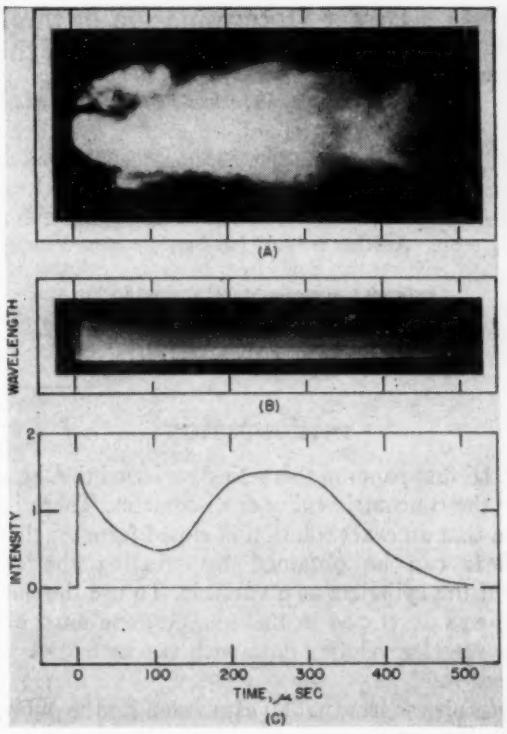


FIG. 5. (a) Discrete photograph of an ultraspeed pellet. (b) Time-resolved spectrogram. (c) Qualitative light curve.

CONCLUSIONS

The representative photographs reproduced herein indicate that AlO bands formed by the ablation and subsequent burning of aluminum from ultraspeed pellets traveling in normal atmospheres seem to be characterized by a delay time of about 150 μsec. This effect has been reproduced about ten times. On the other hand, similar experiments have indicated that the delay time can range as low as 0 μsec. The phenomenon is related to the flashing effect, a space-dependent phenomenon discussed in a previous paper.[6]

ACKNOWLEDGMENT

It is a pleasure to acknowledge the assistance of the members of the group who contributed to the work reported herein. In particular, appreciation is tendered Mr. J. Craig McLaughlin for his ordnance work, and Mr. J. W. Rogers who designed much of the equipment.

[2] Allen and Rinehart, Rev. Sci. Instr. 22, 1020 (1951).
[3] Most of the spectral lines that appear in Fig. 5(b) have been reported in reference 1, p. 200. The presence of Mg and Mn in the commercially pure 2S—O aluminum alloy used for pellet material has been confirmed by quantitative chemical analysis.
[4] Samual Glasstone, editor, *The Effects of Atomic Weapons* (McGraw-Hill Book Company, Inc., New York, 1950), p. 188.
[5] The composition of this alloy is given in reference 1, p. 132.
[6] Rinehart, Allen, and White, J. Appl. Phys. 23, 297 (1952).

Direct Determination of the Flow Curves of Non-Newtonian Fluids. II. Shearing Rate in the Concentric Cylinder Viscometer

IRVIN M. KRIEGER, *Department of Chemistry and Chemical Engineering, Case Institute of Technology, Cleveland, Ohio*
AND
HAROLD ELROD, *Department of Mechanical Engineering, Case Institute of Technology, Cleveland, Ohio*
(Received July 21, 1952)

Another method has been developed for obtaining the rate of shear *vs* shearing stress curves of non-Newtonian fluids from concentric cylinder viscometer data. The mathematical expression developed is a rapidly converging power series in ln$s$, where s is the cup to bob radius ratio. An estimate of error shows that under favorable conditions only two terms of the series are significant, and that terms past the third will hardly ever be needed.

INTRODUCTION

THE first paper in this series[1] reviewed the theory of the concentric cylinder viscometer. There it was shown that an exact solution in closed form for the rate of shear can be obtained by treating the radius ratio of the cylinders as a variable. To use the method which was developed in this manner, one must obtain torque-angular velocity data with two or more bobs of different radii.

It was also shown that an expression for the difference between the rates of shear at the cup and at the bob wall can be recovered from data taken with only one bob. In the present paper, a solution for this difference equation is obtained as a rapidly converging series.

THEORY OF THE CONCENTRIC CYLINDER VISCOMETER

The present treatment presupposes the existence of a functional relationship between the rate of shear dv/dy and the shearing stress F, namely,

$$dv/dy=g(F). \tag{1}$$

The first objective of viscometry is the recovery of this relation, called the flow equation, from experimental data. The flow equation may be expressed in analytic, tabular, or graphical form.

As in the previous paper, we consider a concentric cylinder viscometer having an inner cylinder or bob of length L and radius R_1, and an outer cylinder or cup of radius $R_2=sR_1$. The cup rotates with angular velocity Ω, while an external torque M holds the bob stationary. When the fluid in the annular space between the cup and the bob is in laminar flow, the rate of shear at a distance r from the axis, where the fluid rotates with angular velocity ω, is

$$g(F)=r\frac{d\omega}{dr}=\frac{d\omega}{d\ln r}, \tag{2}$$

while the shearing stress is

$$F=M/2\pi r^2L. \tag{3}$$

Since M is constant under steady flow conditions,

$$d\ln F=-2d\ln r \tag{4}$$

and

$$g(F)=-2(d\omega/d\ln F). \tag{5}$$

At the bob surface, $F=F_1$ and $\omega=0$, while at the cup, $F=F_2$ and $\omega=\Omega$. Integrating (5) between these limits, one obtains

$$\Omega=-\frac{1}{2}\int_{F_1}^{F_2}g(F)d\ln F=-\frac{1}{2}\int_{s^2F_1}^{F_2}\frac{g(F)}{F}dF. \tag{6}$$

In the previous paper, $g(F)$ was obtained by differentiation with respect to s at constant F_2, giving

$$g(F_1)=s(\partial\Omega/\partial s)F_2. \tag{7}$$

From this equation an experimental method was developed for determining the rate of shear from data obtained using two bobs.

When (6) is differentiated with respect to F_1, a difference equation is obtained:[1-3]

$$\frac{d\Omega}{dF_1}=\frac{1}{2F_1}[g(F_1)-g(F_2)]. \tag{8}$$

The method described here for obtaining $g(F)$ is based on a solution of this difference equation.

SOLUTION OF THE DIFFERENCE EQUATION

Let us define a function $h(F_1)$ by the relation

$$h(F_1)=2\frac{d\Omega}{d\ln F_1}. \tag{9}$$

From Eq. (8),

$$h(F_1)=g(F_1)-g(s^{-2}F_1). \tag{10}$$

Also,

$$\begin{aligned}h(s^{-2}F_1)&=g(s^{-2}F_1)-g(s^{-4}F_1),\\ h(s^{-4}F_1)&=g(s^{-4}F_1)-g(s^{-6}F_1),\end{aligned} \tag{11}$$

[1] I. M. Krieger and S. H. Maron, J. Appl. Phys. 23, 147 (1952).
[2] M. D. Hersey, J. Rheol. 3, 196 (1932).
[3] M. Mooney, J. Rheol. 2, 210 (1931).

etc. Since $s>1$ and $g(0)=0$,

$$\sum_{n=0}^{\infty} h(s^{-2n}F_1)=g(F_1). \tag{12}$$

The sum is a slowly convergent one, which may be asymptotically evaluated using the Euler-MacLaurin sum formula,[4]

$$\sum_{n=0}^{m} f(n)=\int_0^m f(x)dx+\tfrac{1}{2}[f(0)+f(m)]+\sum_{k=1}^{r}\frac{B_{2k}}{(2k)!}[f^{(2k-1)}(m)-f^{(2k-1)}(0)]+\sum_{i=0}^{m-1}\int_i^{i+1} P_{2r+1}(x-i)f^{(2r+1)}(x)dx. \tag{13}$$

Here B_i is the ith Bernouilli number and P_i the Bernouilli polynomial. Applying this formula to the summation in Eq. (12),

$$\sum_{n=0}^{\infty} h(s^{-2n}F_1)=\int_0^{\infty} h(s^{-2x}F_1)dx+\tfrac{1}{2}[h(F_1)+h(0)]+\frac{1}{12}\left[\frac{dh(s^{-2n}F_1)}{dn}\right]_{n=0}^{\infty}-\frac{1}{720}\left[\frac{d^3h(s^{-2n}F_1)}{dn^3}\right]_0^{\infty}+\cdots. \tag{14}$$

The integral is readily evaluated by making the substitutions

$$y=s^{-2x}F_1 \tag{15}$$

and

$$dx=-dy/2y\ln s. \tag{16}$$

Therefore, in view of the definition of h,

$$\int_0^{\infty} h(s^{-2x}F_1)dx=\frac{-1}{2\ln s}\int_{F_1}^{0} h(y)d\ln y=\frac{\Omega}{\ln s}. \tag{17}$$

The derivatives appearing in Eq. (14) are

$$\frac{dh(s^{-2n}F_1)}{dn}=-4\ln s\frac{d^2\Omega}{d(\ln s^{-2n}F_1)^2}, \tag{18}$$

$$\frac{d^3h(s^{-2n}F_1)}{dn^3}=-16(\ln s)^3\frac{d^4\Omega}{d(\ln s^{-2n}F_1)^4}, \tag{19}$$

etc.

Since $h(F_1)$ and its derivatives are zero when $F_1=0$, the final asymptotic expression for the rate of shear may be written

$$g(F_1)=\frac{\Omega}{\ln s}\left[1+\ln s\cdot\frac{d\ln\Omega}{d\ln F_1}+\frac{(\ln s)^2}{3\Omega}\cdot\frac{d^2\Omega}{d(\ln F_1)^2}-\frac{(\ln s)^4}{45\Omega}\cdot\frac{d^4\Omega}{d(\ln F_1)^4}+\cdots\right]. \tag{20}$$

[4] J. Jeffreys and B. S. Jeffreys, *Methods of Mathematical Physics* (Cambridge University Press, Cambridge, 1946), p. 255.

A particularly attractive feature of the above result is the fact that the bracketed term is a power series in $\ln s$. The better concentric cylinder viscometers are constructed with small clearances, so that $\ln s$ rarely exceeds 0.2 and is frequently 0.05 or less. The term $d\ln\Omega/d\ln F_1$ is unity for Newtonian fluids, but may take on higher values for non-Newtonian fluids. (Values as high as five have been observed in this laboratory.) The second term is therefore significant.

To estimate the error involved in terminating the series at any given point, it is convenient to consider the special case of a fluid whose flow equation is

$$g(F)=aF^N, \tag{21}$$

where a and N are constants characteristic of the fluid.[5-7] Integration of Eq. (6) with this value of $g(F)$ gives

$$\Omega=a/2N(1-s^{-2N})F_1^N; \tag{22}$$

hence

$$d\ln\Omega/d\ln F_1=N,$$
$$d^3\Omega/d(\ln F_1)^3=N^3\Omega, \tag{23}$$

etc. Thus

$$g(F_1)=\frac{\Omega}{\ln s}\left[1+N\ln s+\frac{(N\ln s)^2}{3}-\frac{(N\ln s)^4}{45}+\cdots\right]. \tag{24}$$

Hence the error in terminating the series at the second term is of order $\frac{1}{3}(N\ln s)^2$, while carrying the third term reduces the error to $1/45(N\ln s)^4$.

In this connection, it is evident that an instrument for which s is small is desirable. Comparison of two common concentric cylinder viscometers brings out this point. For the Precision-Interchemical viscometer,[8] $R_2=1.500$ cm, $R_1=1.300$ cm, making $s=1.154$. Here N must be less than 1.4 in order that the series may be terminated at two terms without exceeding 1 percent error, while N values up to 7.0 may be used if the third term is carried. For the Mooney-Ewart coni-cylindrical viscometer,[9] for which $R_1=2.00$, $R_2=2.10$, and $s=1.05$, the corresponding N values are 4.1 for two terms and 20.5 for three. The computational problem is thus much simpler in the latter case, since a second differentiation will rarely be required.

USE OF THE METHOD

In practice, M *vs* Ω data are obtained, and F_1 calculated by multiplying M by the factor $1/2\pi R_1^2L$. A plot of $\log\Omega$ *vs* $\log F_1$ is constructed. If this plot is linear, the exponential flow equation applies, and $g(F)$ may be obtained by well-established techniques.[7] If this graph is curved, graphical differentiation at points along the

[5] A. W. Porter and P. A. M. Rao, Trans. Faraday Soc. **23**, 311 (1927).
[6] Farrow, Lowe, and Neale, J. Textile Inst. **19**, T 18 (1928).
[7] I. M. Krieger and S. H. Maron, J. Colloid Sci. **6**, 528 (1952).
[8] H. Green, *Industrial Rheology* (John Wiley and Sons, Inc., New York, 1949), p. 100.
[9] M. Mooney and R. H. Ewart, J. Appl. Phys. **5**, 530 (1934).

curve yields the slope m, where

$$m = d\ln\Omega/d\ln F_1. \tag{25}$$

For those cases where $m\ln s$ is less than 0.2, the third and higher terms may be neglected, and the rate of shear calculated with an error of less than 1 percent by

$$g(F_1) = \Omega/\ln s(1+m\ln s). \tag{26}$$

When $m\ln s$ is greater than 0.2 but less than 1, an additional graphical differentiation is required to keep the accuracy within 1 percent. Since

$$\frac{1}{\Omega}\frac{d^2\Omega}{d(\ln F_1)^2} = m^2 + \frac{dm}{d\ln F_1}, \tag{27}$$

the third term may be obtained from slopes of a graph of m *vs* $\ln F_1$. The rate of shear is then

$$g(F_1) = \frac{\Omega}{\ln s}\left[1 + m\ln s + \tfrac{1}{3}(m\ln s)^2 + \frac{(\ln s)^2}{3}\frac{dm}{d\ln F_1}\right]. \tag{28}$$

In the rare case when $m\ln s$ is greater than 1, the fourth term may be required, although this has never been necessary for the systems studied in this laboratory. For most fluids, the first two terms should suffice.

CONCLUSIONS

An alternate method has been devised for the recovery of the rate of shear of a non-Newtonian fluid from data obtained in concentric cylinder viscometers, thus permitting the direct determination of the flow curves of such fluids without prior assumption of a flow equation. The mathematical expression for the rate of shear is a power series in the logarithm of the radius ratio; the coefficients are derivatives of the angular velocity with respect to shearing stress. Under favorable conditions, terms beyond the second are negligible, while inclusion of the third term is almost always adequate. Instruments designed with radius ratios near unity are more satisfactory in this respect, since only one differential need be evaluated.

A paper is in preparation in which the flow curves of several fluids, obtained by the various direct methods, are intercompared. The work discussed herein was performed as part of a research project sponsored by the Reconstruction Finance Corporation, Office of Synthetic Rubber, in connection with the Government Synthetic Rubber program. The advice and encouragement of Dr. Samuel H. Maron are acknowledged.

JOURNAL OF APPLIED PHYSICS VOLUME 24, NUMBER 2 FEBRUARY, 1953

The Nature of the Coefficient of Friction

J. T. Burwell * and E. Rabinowicz
Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received August 18, 1952)

The modern theory of the friction between dry metal surfaces ascribes it to local minute welds or adhesions between the surfaces and suggests that for a given pair of surfaces the friction force is uniquely defined by the normal load alone. Herein it is demonstrated that this cannot in general be true and that some further condition of operation must also be defined. Experiments are reported indicating that one such possible condition is the sliding speed so that the friction force is actually a function of the normal load and the sliding speed. It is pointed out that the speed can influence the friction force in two ways—one, by the resulting shear strain rate in the vicinity of the welded junction, and the other by the length of time taken for a junction of full strength to form.

THE current theory of the mechanism of dry friction[1-3] pictures the force of friction as arising from the shear strength of minute welds formed between two bodies in contact and distributed more or less at random over their apparent contact area. This can be expressed by the relation for the friction force

$$F = s\cdot A, \tag{1}$$

where s is the average shear strength of these welds and A the sum of their individual areas. Since it is these areas that actually carry the normal load W between the two bodies, we also have

$$W = p_m\cdot A, \tag{2}$$

where p_m is defined as the flow pressure of the softer material in the vicinity of these local true contact areas. Eliminating A between these two equations leads to the familiar expression for the friction coefficient

$$\mu = F/W = s/p_m. \tag{3}$$

Bowden and Tabor emphasize the point that this is the ratio of two plastic properties of the weld and adjacent material. It has been customarily assumed that these two quantities are each constant for a given pair of contacting surfaces so that μ is also a constant.

Recently, however, McFarlane and Tabor[4] have

* Now with Horizons, Inc., Cleveland, Ohio.

[1] R. Holm, *Electric Contacts* (Almquist and Wiksells, Uppsala, 1946).

H. Ernst and M. E. Merchant, Proc. M.I.T. Sum. Conf. on Surface Finish, p. 76, (June 1940).

[3] F. P. Bowden and D. Tabor, *The Friction and Lubrication of Solids* (Oxford University Press, New York, 1950).

[4] J. S. MacFarlane and D. Tabor, Proc. Roy. Soc. (London) A202, 244 (1950).

pointed out that, for materials that show plasticity, the shear stress and normal stress must be related by a condition of plasticity. Using a general form of the von Mises criterion, they were able to explain certain adhesion behavior between a steel and an indium surface, and specifically, that the friction coefficient was, in fact, not a constant independent of the load. Independently, Parker and Hatch[5] have observed visually the growth of the true contact area between indium and glass when a tangential force is applied between the two. This growth can also be explained very easily by applying a criterion of the von Mises type. Hence, it would appear that to the system to which Eqs. (1) and (2) apply, there should be added a condition for plastic flow which, in its most general form, may be written

$$f(s,p_m,Y)=0, \tag{4}$$

where Y is the yield strength in tension of the material in question and is an invariant property, but s and p_m are no longer constant. This would be true, regardless of whether the shear takes place at the interface of the two materials or within the softer one.

Consider the situation of a purely plastic material pressed against another considerably harder surface under a given normal load to which a tangential force is then applied to maintain continued sliding. It may then be asked what is the value of this friction force. Equations (1), (2), and (4) apply and contain the three dependent quantities s, p_m, and A. Hence they are not sufficient to determine F. A search for additional information does not appear to yield a determinate problem.†

This, therefore, suggests that in such a sliding system the friction force is not determinate. Presumably some other conditions must be imposed, which does, in fact, apply in actual sliding. For a plastic material one

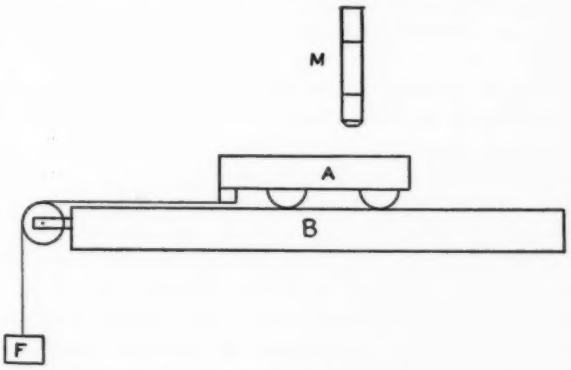


FIG. 1. Schematic diagram of friction apparatus.

[5] R. C. Parker, and D. Hatch, Proc. Phys. Soc. (London) B63, 185 (1950).

† It is true that a definite geometry, say having circular symmetry, can be assumed for a local contact region so that the equations of equilibrium and the Levy-Mises relations can be used. This does not, in fact, introduce any additional relations between s, p_m, and A, so that the number of relations is still not sufficient to determine the friction force.

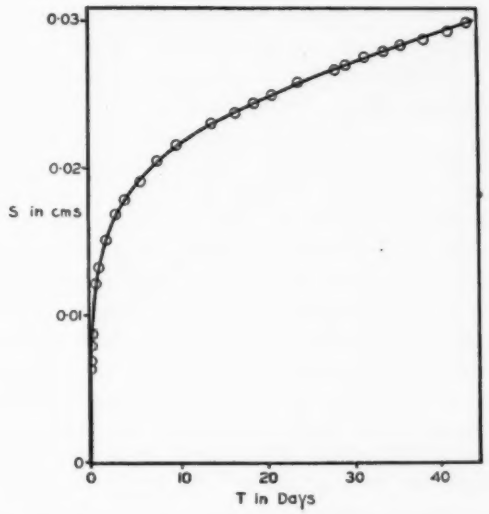


FIG. 2. Displacement as a function of time. Steel on indium, clean; load 300 g, tangential force 200 g.

obvious additional condition is the dependence of shear stress on the rate of shear, i.e., the sliding velocity. This possibility could be tested by measuring the friction force at a number of different positively imposed sliding speeds to see whether there is a unique dependence between the two. A simpler method of obtaining the same information, however, is to impose a constant tangential force as by means of a dead weight and to measure the resulting velocity. To test this idea some simple friction experiments have been carried out using a very plastic material such as indium or lead as one of the two surfaces.

The apparatus used is shown schematically in Fig. 1. A brass block A into one surface of which had been screwed three steel hemispherically ended sliders of diameter 8 mm, rested on a clean, flat, horizontal surface B, consisting of the soft metal, indium or lead. A tangential force was applied by a load F and pulley arrangement, the point of application of the force being about 1 mm above the indium surface. The total weight of the slider was 300 g. The displacement was measured by observations of the movement of a scratch in the upper surface of the brass block in relation to the graduated scale of a microscope M fixed with respect to the bottom surface. In the first experiments the bottom surface was indium.

According to previous theories, the results to be expected in such an experiment should be a small rapid displacement, which is due partly to the elastic forces in the steel and indium (see Rankin[6]) and partly to the interaction of the normal and tangential stress in a plastic material as postulated by McFarlane and Tabor[4] and observed by Parker and Hatch.[5] This should be almost instantaneous, and thereafter all movement should cease. The actual curve of displacement as a function of time for a typical case, force = 200 g

[6] J. S. Rankin, Phil. Mag. 2, 806 (1926).

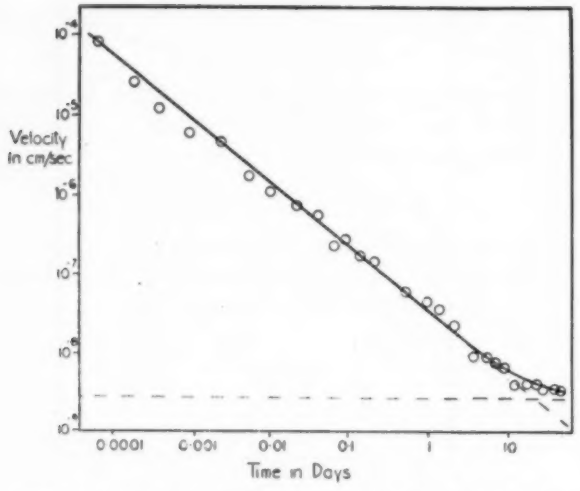


FIG. 3. Sliding velocity as a function of time for the data shown in Fig. 2.

(i.e., $\mu=0.67$), is shown in Fig. 2, and it will be observed that continuous movement has taken place.

It was found that the experimental results obtained were almost unaffected by the occurrence of vibration in the laboratory, and this stability was probably due to the large normal adhesion of steel to indium.[4]

Investigation of the curve of Fig. 2 showed that it fitted a relation of the type

$$s=6\cdot C_1\cdot T^{\frac{1}{2}}+C_2\cdot T, \tag{5}$$

$$\text{whence } v=C_1\cdot T^{-\frac{1}{2}}+C_2. \tag{6}$$

Figure 3 shows a plot of velocity against time on a logarithmic scale for the data shown in Fig. 1, and it will be seen that for the first ten days of the experiment the first term of Eq. (6) was predominant; thereafter the second term became more important. The final constant velocity is about $2\cdot10^{-9}$ cm/sec, or $2\cdot10^{-4}$ cm/day, and may be considered a true kinetic velocity for steel sliding on indium.

By applying different tangential forces, a steady state of sliding was obtained in each case after an initial transient stage, the steady state being reached more quickly at high than low values of μ. The final velocities were different for each tangential force until a critical value was reached above which no steady sliding state could be attained, the block accelerating continuously on application of the constant tangential force. This corresponds to a maximum value of the friction coefficient μ. A plot of friction coefficient against steady-state velocity is shown in Fig. 4, which shows μ_{max} and also contains a number of points obtained by measuring the coefficient of friction at relatively much higher imposed sliding velocities using conventional methods.

Similar results were obtained with a somewhat harder metal, lead, although the rates were a smaller order of magnitude. The resulting curve of μ *vs* V is shown

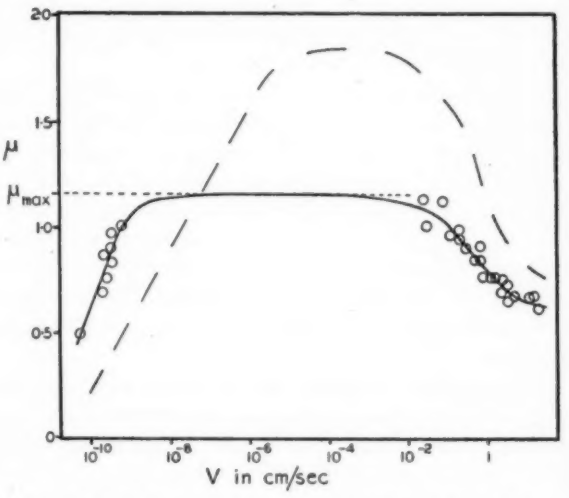


FIG. 5. Coefficient of friction as a function of final sliding velocity. Clean steel on lead, load 300 g.

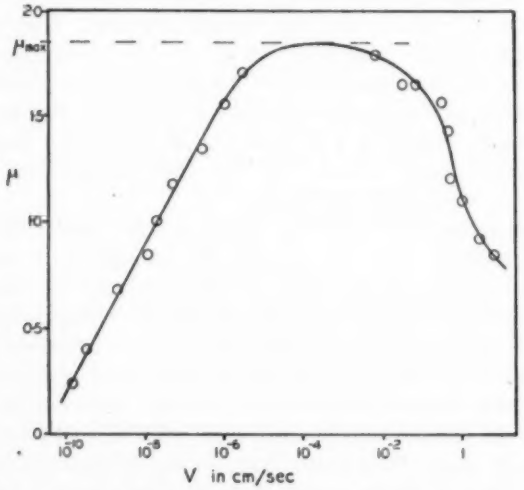


FIG. 4. Coefficient of friction as a function of final sliding velocity. Clean steel on indium, load 300 g.

in Fig. 5 where the corresponding curve for indium from Fig. 4 is shown dashed for comparison. It is to be noted that the value of μ_{max} for the lead is considerably less than for the indium. However, for a given sliding speed to the left of the maximum μ for the lead is higher.

A curve similar to those of Figs. 4 and 5 has been observed by Papenhuyzen[7] for rubber, and Bristow[8] has suggested that metals also show relative motion at tangential forces insufficient to produce large-scale sliding. Furthermore, common experience with a number of other nonmetals such as the plastics on wood and glass suggest that they show a similar effect.

The curves in Figs. 4 and 5 can also be looked at from a somewhat different viewpoint. They represent, of course, the creep behavior of the system in shear. Points

[7] P. J. Papenhuysen, D. Ingenieur **53**, 75 (1938),
[8] J. R. Bristow, Proc. Roy. Soc. (London) A189, 88 (1947).

to the left of the maximum show only the primary and secondary phases of creep as evidenced by Fig. 3 and Eq. (6). From the point of view of creep testing, the present system is unusual in that its geometry allows the test to proceed indefinitely without the possibility of failure through necking, i.e., cross-sectional area is continually being added on the forward side of each weld as sliding proceeds to compensate for that lost on the trailing side, thus remaining constant. The primary portion of the velocity-time curve in Fig. 3 varies as the $-\frac{5}{6}$ power of the time. This is intermediate between the $-\frac{2}{3}$ and -1 power of the time, both of which have experimental as well as theoretical support. Recently Wyatt[9], conducting creep tests under tension, has also observed this intermediate behavior. The straight-line relationship between applied stress and the logarithm of the final strain rate as shown by the left-hand portion of the curves of Figs. 4 and 5 is also in agreement with the predictions of theory, but does not occur often in practice.[10]

Although points to the left of the maximum of the curves in Figs. 4 and 5 do not show failure, points to the right of the maximum do. This failure, in the general case, could be either ductile or brittle. Probably near the maximum the failure is ductile and results from necking or reduction in cross-sectional area of the welds. Recent results of Rabinowicz,[11] showing how at moderate velocities the friction force varies with displacement from rest, are consistent with this idea.

The reason that such necking takes place here and not at lower speeds is as follows. There is considerable evidence that it takes a finite time for a weld, when formed, to reach its full strength. This is seen more clearly in friction experiments of the "stick-slip" type where the length of time that the two surfaces are at rest can be most conveniently varied over the range of interest by merely varying the nominal sliding speed of the surfaces. In such cases the static or breakaway friction decreases as this time decreases.[12] In the present case, as the speed increases toward the maximum of the curves in Figs. 4 and 5, the time of contact begins to become insufficient for the region on the forward side of the weld to grow to its full strength. Finally, to the right of the maximum this weakening outweighs the increased force required by the increased strain rate. Hence, failure in this region results through decrease in the effective cross-sectional area. This situation is unstable, resulting in the "stick-slip", chattering, and sqeaking commonly observed in dry friction.

At somewhat higher sliding speeds the fracture may become brittle in certain materials. Its occurrence will depend on the relation of the brittle strength to the yield stress in tension[13] of the material in question, on the speed and on the temperature.

With considerably harder metals such as aluminum, copper, steel, etc., the maximum of the curve such as in Fig. 4 or 5 probably lies so near the ordinate of zero velocity that the left-hand portion of the curve cannot be detected in a reasonable length of time, and only the right-hand descending portion is ever observed.

From the above discussion it will be seen that the curves of Figs. 4 and 5 are inconsistent with the concept of a unique value of μ, since any value of the friction force (so long as $\mu < \mu_{max}$) can be imposed on the system and it will always find a velocity v at which movement will take place. Hence, in the system of Eqs. (1), (2), and (4) F is independently determined by the conditions of the experiment, and the equations are then sufficient to determine the values of shear strength, flow pressure, and true contact area. Also, the values of s, p_m, and A will not be constant for a given pair of metals but will vary depending on the applied tangential force. This force may be varied directly as in the present experiments, thus yielding different sliding speeds, or conversely, the sliding speed may be varied, and varying tangential forces will result. At the higher speeds the additional factor of the time required to form a full-strength weld also enters, so that again the friction coefficient depends on the sliding speed.

The above is the situation once sliding has begun. If the surfaces are initially at rest then this analysis indicates that a varying value of the so-called static friction coefficient will be observed depending on what sliding speed one attempts to initiate. This would ordinarily be determined by the rate of application of the tangential force. In particular, Fig. 4 shows that for steel on indium any value up to 1.8 or even higher can be obtained with care depending on (a) the rate of loading, (b) the time taken to observe whether sliding has actually taken place, and (c) the transient creep phenomenon. Add to these factors that of the length of time prior to contact, and it is no wonder that such a range of values of static friction coefficient appear in the literature, even for reproducibly cleaned metal surfaces.

ACKNOWLEDGMENTS

The authors are indebted to the Chrysler Corporation whose generous support made this research possible, and to the Indium Corporation of America for the gift of the indium metal.

[9] O. H. Wyatt, Nature **167**, 866 (1951).
[10] I. S. Servi and N. J. Grant, J. Inst. Metals **80**, 33 (1951).
[11] E. Rabinowicz, J. Appl. Phys. **22**, 1373 (1951).
[12] Sampson, Morgan, Reed, and Muskat, J. Appl. Phys. **14**, 689 (1943).
[13] E. Orowan, Reports on Progress in Physics (The Physical Society of London) **12**, 185 (1949).

An Alternative Method of Solving Hallén's Integral Equation and Its Application to Antennas Near Resonance

RONOLD KING
Cruft Laboratory, Harvard University, Cambridge, Massachusetts
(Received September 26, 1952)

Hallén's complex integral equation for the current in a cylindrical antenna is separated into two real integral equations for the components of current in phase and in phase-quadrature with the driving voltage. Each of these equations is solved by iteration using zeroth-order currents and vector-potential differences to define expansion parameters. It is shown that for electrical half-lengths near odd multiples of a quarter-wavelength at least a third-order solution is required in order to determine accurately the component of current in phase with the driving voltage and the conductance. Conductances for a range of radii are evaluated by the new third-order formula and compared with the King-Middleton second-order values and with the experimental data of Hartig. The new formula agrees excellently with experimental results at $h=\lambda_0/4$, whereas the earlier second-order formula has by far its greatest percent error—near 8 to 10 percent—in a small range near resonance. It is concluded that for antennas near resonance, just as for very short and very long antennas, adequate account must be taken in the iteration of both components of current.

INTRODUCTION

IT is shown in a recent paper[1] that a more accurate determination of the circuit properties of an electrically short antenna may be obtained from the integral equation of Hallén,[2]

$$\int_0^h I(z')\left[\frac{e^{-j\beta_0 R_1}}{R_1}+\frac{e^{-j\beta_0 R_2}}{R_2}\right]dz' = \frac{-j4\pi}{\zeta_0}\left[C\cos\beta_0 z+\tfrac{1}{2}V_0\sin\beta_0 z\right], \quad (1)$$

if the unknown complex current

$$I(z')=I_r(z')+jI_i(z') \quad (2)$$

is inserted in (1) and this is separated into the following two real integral equations:

$$\int_0^h [I_r(z')K_c(z,z')-I_i(z')K_s(z,z')]dz' = \frac{4\pi}{\zeta_0}C'F_0(z), \quad (3a)$$

$$\int_0^h [I_i(z')K_c(z,z')+I_r(z')K_s(z,z')]dz' = -\frac{4\pi}{\zeta_0}[C'F_0(z)+\tfrac{1}{2}V_0G_0(z)]. \quad (3b)$$

In (1) through (3b) the following notation is used:

$$R_1=[(z-z')^2+a^2]^{\frac{1}{2}}; \quad R_2=[(z+z')^2+a^2]^{\frac{1}{2}}, \quad (3c)$$

$$K_c(z,z')=\frac{\cos\beta_0 R_1}{R_1}+\frac{\cos\beta_0 R_2}{R_2};$$

$$K_s(z,z')=-\left[\frac{\sin\beta_0 R_1}{R_1}+\frac{\sin\beta_0 R_2}{R_2}\right]. \quad (3d)$$

$C=C''+jC'$, $\beta_0=2\pi/\lambda_0$, $\zeta_0=120\pi$ ohms, V_0 is the driving voltage at the center of the antenna in the form of a discontinuity in scalar potential. (Currents in this idealized case correspond to the extrapolated values at zero line spacing of currents measured over a range of line spacings.) h is the half-length; a, the radius of the cylindrical antenna. It is assumed that $a\ll h$, $\beta_0 a\ll 1$. Also

$$F_0(z)\equiv\cos\beta_0 z; \quad G_0(z)\equiv\sin\beta_0 z. \quad (3e)$$

The reason why (3a) and (3b) are preferred to (1) under certain conditions arises from the fact that in carrying out the iteration of (1) according to the King-Middleton procedure[3,4] the zeroth-order solution is the leading term in the quadrature current $I_i(z)$. It follows that the in-phase current $I_r(z)$ does not occur until first-order terms are evaluated, so that it is always one step lower in the iteration. Since $I_i(z)$ is by far the more important component in maintaining the far-zone field for a wide range of practically important lengths of antenna, this situation is entirely satisfactory except under special circumstances. These include the following: (a) As explained in detail in reference 1, the component $I_r(z)$ is so small compared with $I_i(z)$ in an electrically short antenna that quantities such as the resistance and the gain which depend upon $I_r(z)$ are not determined accurately in the first two iterations of (1), using the zeroth-order current to define the expansion parameter. (b) For antennas near resonance, i.e., $\beta_0 h$ near $n\pi/2$, n odd, the component $I_i(z)$ all but vanishes while $I_r(z)$ becomes large. (c) In the case of very long antennas with $\beta_0 h>2\pi$, the component $I_r(z)$ is comparable in magnitude with $I_i(z)$ for all values of $\beta_0 h$.

It is the purpose of this paper to solve (3a, b) in general for use in a more accurate solution in cases (b) and (c) and to apply this solution specifically and in detail to case (b). A detailed study of the long antenna, case (c), is reserved for another report.

[1] R. King, "Theory of electrically short transmitting and receiving antennas," Cruft Laboratory Technical Report No. 141 (March, 1952); J. Appl. Phys. 23, 1174 (1952).
[2] E. Hallén, Nova Acta Roy. Soc. Sci., Upsala 11, 1 (1938).
[3] R. King and D. Middleton, Quart. Appl. Math. 3, 302 (1946).
[4] D. Middleton and R. King, J. Appl. Phys. 17, 273 (1946).

SOLUTION OF THE INTEGRAL EQUATIONS FOR THE COMPONENTS OF CURRENT

Instead of solving (3a) and (3b) using the kernels $K_c(z,z')$ and $K_s(z,z')$ defined in (3d), it is advantageous to introduce the difference kernels of the type found so useful in the analysis of the short antenna.[1] These are obtained by subtracting, respectively, from (3a) and (3b), the same equations with $z=h$. This gives

$$\int_0^h [I_r(z')L_c(z,z')-I_t(z')L_s(z,z')]dz' = \frac{4\pi}{\zeta_0}C'F_{0z}, \quad (4a)$$

$$\int_0^h [I_t(z')L_c(z,z')+I_r(z')L_s(z,z')]dz' = \frac{-4\pi}{\zeta_0}[C''F_{0z}+\tfrac{1}{2}V_0G_{0z}], \quad (4b)$$

where

$$L_c(z,z')=K_c(z,z')-K_c(h,z'), \quad (4c)$$

$$L_s(z,z')=K_s(z,z')-K_s(h,z'), \quad (4d)$$

$$F_{0z}=F_0(z)-F_0(h), \quad (4e)$$

$$G_{0z}=G_0(z)-G_0(h). \quad (4f)$$

The advantage in using the difference kernels may be explained as follows. Since $K_s(z,z')$ reduces to extremely small values at $z'=z$, whereas $K_c(z,z')$ has its maximum value at $z'=z$, it follows that

$$\int_0^h I(z')K_s(z,z')dz'$$

is relatively small and slowly varying as a function of z, whereas

$$\int_0^h I(z')K_c(z,z')dz'$$

varies approximately with the current and thus assumes values at points along the antenna that greatly exceed the small value at its end, $z=h$. Accordingly, an integral involving the difference kernal $L_s(z,z')$ is relatively much smaller compared with an integral involving $L_c(z,z')$ than when $K_s(z,z')$ and $K_c(z,z')$ are used in their places. An approximate solution of (4a, b) for the components of current, therefore, is more accurate than a corresponding solution of (3a, b). In particular, the solution of (4a) as an equation in $I_r(z)$ with a less essential contribution from $I_t(z)$, and of (4b) as an equation in $I_t(z)$ with only a small contribution from $I_r(z)$, is suggested.

After the introduction of a distribution function for each component of current in the manner of reference 3 or 4, i.e.,

$$g_j(z,z')=I_j(z')/I_j(z); \quad j=r \text{ or } t \quad (5)$$

and the definition of

$$W_j(z)=W_j+\sigma_j(z)\equiv\int_0^h g_j(z,z')L_c(z,z')dz'; \quad j=r \text{ or } t, \quad (6)$$

where W_j is the sensibly constant value of $W_j(z)$ at an appropriate reference point $z=z_r$ and $\sigma_j(z)$ is a correction term that is very small for most values of z, the following equations are obtained:

$$I_r(z)=\frac{4\pi}{\zeta_0 W_r}C'F_{0z}+\frac{1}{W_r}[D_r(z)-E_t(z)], \quad (7a)$$

$$I_t(z)=-\frac{4\pi}{\zeta_0 W_t}[C''F_{0z}+\tfrac{1}{2}V_0G_{0z}] + \frac{1}{W_t}[D_t(z)+E_r(z)], \quad (7b)$$

where the difference integrals are

$$D_j(z)\equiv -I_j(z)\sigma_j(z)-\int_0^h [I_j(z')-I_j(z)g_j(z,z')]L_c(z,z')dz'; \quad j=r \text{ or } t \quad (8a)$$

$$E(z)\equiv -\int_0^h I_j(z')L_s(z,z')dz'; \quad j=r \text{ or } t. \quad (8b)$$

Solutions of (7a) and (7b) may be derived using the method of iteration described in reference 3 or 4 with zeroth-order solutions given by

$$[I_r(z)]_0=\frac{4\pi}{\zeta_0 W_r}C'F_{0z}, \quad (9a)$$

$$[I_t(z)]_0=-\frac{4\pi}{\zeta_0 W_t}[C''F_{0z}+\tfrac{1}{2}V_0G_{0z}]. \quad (9b)$$

The corresponding expressions, after a first iteration, are

$$[I_r(z)]_1=\frac{4\pi}{\zeta_0 W_r}\{C'[F_{0z}+Fc_{1z,r}W_r^{-1}] + W_t^{-1}[C''Fs_{1z}+\tfrac{1}{2}V_0Gs_{1z}]\}, \quad (10a)$$

$$[I_t(z)]_1=-\frac{4\pi}{\zeta_0 W_t}\{C''[F_{0z}+Fc_{1z,t}W_t^{-1}] + \tfrac{1}{2}V_0[G_{0z}+G_{1z,t}W_t^{-1}]-C'Fs_{1z}W_r^{-1}\}, \quad (10b)$$

where Fc_{1zj} and Gc_{1zj}, $j=r$ or t, are given, respectively, by (8b) with F_{0z} and G_{0z} substituted for $I_j(z)$, and Fs_{1z} and Gs_{1z} are given by (8b) with F_{0z} and G_{0z} substituted for $I_j(z)$. Higher order terms may be obtained by continuing the iteration. Note that Fc_{1z} and Fs_{1z} may be separated so that $Fc_{1z}=Fc_1(z)-Fc_1(h)$,

$Fs_{1z}=Fs_1(z)-Fs_1(h)$, where $Fc_1(z)$ and $Fs_1(z)$ are given by (8a) and (8b), respectively, with $K_c(z,z')$ and $K_s(z,z')$ substituted for $L_c(z,z')$ and $L_s(z,z')$. When $z=h$ these values reduce to

$$Fc_1(h)\equiv Fc_1=-\int_0^h F_{0z'}K_c(h,z')dz', \tag{11a}$$

$$Fs_1(h)\equiv Fs_1=-\int_0^h F_{0z'}K_s(h,z')dz'. \tag{11b}$$

Similar expressions are obtained by replacing F by G.

The constants C' and C'' may be evaluated from (3a, b) by setting* $z=h$ and using the curretns (9a, b) for first-order values, (10a, b) for second-order values, etc. The first-order values are

$$C'=\tfrac{1}{2}V_0N'/W_0\Delta;\quad C''=-\tfrac{1}{2}V_0N''/\Delta, \tag{12a}$$

where

$$N'=[Fs_1(G_0+Gc_1W_t^{-1})-Gs_1(F_0+Fc_1W_t^{-1})], \tag{12b}$$

$$N''=[(F_0+Fc_1W_r^{-1})(G_0+Gc_1W_t^{-1})+Fs_1W_r^{-1}Gs_1W_t^{-1}], \tag{12c}$$

$$\Delta=(F_0+Fc_1W_r^{-1})(F_0+Fc_1W_t^{-1})+Fs_1^2W_r^{-1}W_t^{-1}. \tag{12d}$$

With (12a) through (12d), 10a) and (10b) may be expanded into the following expressions:

$$I_r(z)=\frac{2\pi V_0}{\zeta_0W_rW_tD_r}\{\{F_0(z)[G_0Fs_1-F_0Gs_1]-F_0[Fs_1(z)G_0-Gs_1(z)F_0]\}+W_r^{-1}\{Fc_{1p}(z)[G_0Fs_1-F_0Gs_1]$$
$$-Fc_1[Fs_1(z)G_0-Gs_1(z)F_0]\}+W_t^{-1}\{F_0(z)[Fs_1Gc_1-Gs_1Fc_1]-F_0[Fs_1(z)Gc_1-Gs_1(z)Fc_1]\}$$
$$+W_r^{-1}W_t^{-1}\{\cdots\}+\cdots\}; \tag{13a}$$

$$I_t(z)=\frac{2\pi V_0}{\zeta_0W_t}\left\{\frac{\sin\beta_0(h-z)+W_t^{-1}M_{1t}{}'(z)+W_t^{-2}\{\cdots\}+\cdots}{F_0+W_t^{-1}Fc_1+W_r^{-1}W_t^{-1}Fs_1^2[F_0+Fc_1W_t^{-1}]^{-1}}+\frac{W_r^{-1}W_t^{-1}\{\cdots\}+\cdots}{D_r}\right\}, \tag{13b}$$

where

$$D_r\equiv F_0^2+(W_r^{-1}+W_t^{-1})F_0Fc_1+W_r^{-1}W_t^{-1}(Fc_1^2+Fs_1^2+\cdots)+\cdots \tag{13c}$$

and

$$M_t{}^I(z)\equiv F_0(z)Gc_1-G_0(z)Fc_1+Fc_{1q}(z)G_0-Gc_{1q}(z)F_0. \tag{13d}$$

A comparison of (13b) with the imaginary part of the King-Middleton formula [(23) in reference 3] shows that the zeroth- and first-order terms are, in fact, identical except for the appearance of W_t as expansion parameter instead of Ψ, since difference kernels are used. It may be concluded that the solution for the quadrature current $I_t(z)=I_z'$, obtained previously in (23) of reference 3, is entirely adequate.

The component of current in phase with the driving voltage is seen to have a leading term that is essentially of first-order so that the simpler King-Middleton method of iteration[3,4] (in which the zeroth-order current used in the iteration of the complex integral equation (1) is the leading term of the quadrature current) is justified. Moreover, the leading term in the numerator of (13a) differs relatively little from the leading real term $M_1{}^{II}(z)$ in (23) of reference 3. However, when β_0h is sufficiently near an odd multiple of $\pi/2$, the zeroth- and first-order terms in the denominator of (13a) become small or vanish so that the remaining leading terms actually are of second order. Specifically, when $\beta_0h=n\pi/2$, n odd, the zeroth-order distributions of current for $I_r(z)$ and $I_t(z)$ are alike so that $W_r=W_t=W$ with $g(z,z')$ in (5) equal to $\cos\beta_0z'/\cos\beta_0z$. It follows that (13a) reduces to

$$I_r(z)=\frac{2\pi V_0}{\zeta_0}\left\{\frac{[Fs_1+W^{-1}(Fs_1Gc_1-Gs_1Fc_1)]\cos\beta_0z+W^{-1}[Fc_1(z)Fs_1-Fs_1(z)Fc_1]+\cdots}{Fc_1^2+Fs_1^2+2W^{-1}[Fc_1(Fc_s-Fss_s)+Fs_1(Fs_s+Fcs_s)]}\right\}, \tag{14}$$

where $(Fc_1^2+Fs_1^2)$ in (14) is the complete second-order term and the rest of the denominator is the complete third-order term when $\beta_0h=n\pi/2$, n odd. This latter is obtained by substituting (10a, b) in (3a, b) in order to evaluate Δ in (12d) to a higher order. Since any contributions to the third-order value of Δ, obtained by substituting third-order currents in (3a, b), must have $F_0=\cos\beta_0h$ as a factor, they evidently reduce to zero, when $\beta_0h=n\pi/2$, $n=$ odd.

The following formulas are valid when $\beta_0h=n\pi/2$, n odd:

$$F_{0z}=F_0(z)-F_0(h)=\cos\beta_0z;$$

$$Fc_{1z}=Fc_1(z)-Fc_1(h), \tag{15}$$

$$Fc_1(z)=W\cos\beta_0z-\int_0^h K_c(z,z')\cos\beta_0z'dz'$$
$$=W\cos\beta_0z-\text{R.P.}C_a(h,z), \tag{16a}$$

* Any other convenient value of z may be used. In particular, for very short antennas $z=0$ is appropriate as shown in reference 1. For long antennas, $z=h-\lambda_0/4$ may be desirable.

$$Fs_1(z) = -\int_0^h K_s(z,z')\cos\beta_0 z'dz' = -\text{I.P.}C_a(h,z), \quad (16b)$$

$$Gc_1(z) = W\sin\beta_0 z - \int_0^h K_c(z,z')\sin\beta_0 z'dz'$$

$$= W(\sin\beta_0 z - 1) - \text{R.P.}[S_a(h,z) - E_a(h,z)], \quad (16c)$$

$$Gs_1(z) = -\int_0^h K_s(z,z')\sin\beta_0 z'dz'$$

$$= -\text{I.P.}[S_a(h,z) - E_a(h,z)]. \quad (16d)$$

By setting $z=h$ in (16a)–(16d) the following are obtained:

$$Fc_1 = Fc_1(h) = -\text{R.P.}C_a(h,h) \doteq -\tfrac{1}{2}\,\text{Si}4\beta_0 h, \quad (17a)$$

$$Fs_1 = Fs_1(h) = -\text{I.P.}C_a(h,h) \doteq +\tfrac{1}{2}\,\text{Cin}4\beta_0 h, \quad (17b)$$

$$Gc_1 = Gc_1(h) = -\text{R.P.}[S_a(h,h) - E_a(h,h)]$$
$$\doteq -\tfrac{1}{2}\,\text{Cin}4\beta_0 h + 2\ln 2, \quad (17c)$$

$$Gs_1 = Gs_1(h) = -\text{I.P.}[S_a(h,h) - E_a(h,h)]$$
$$\doteq -\tfrac{1}{2}\,\text{Si}4\beta_0 h. \quad (17d)$$

TABLE I. Conductance and resistance at $\beta_0 h = \pi/2$.[a]

Ω $(=2\ln 2h/a)$	W	$[G_0]_1$	$[G_0]_2$	$[R_0]_1$	$[R_0]_2$
7	4.74	8.35×10^{-3}	9.16×10^{-3}	107.9	95.7
8	5.70	8.73	9.24	99.3	91.6
9	6.68	8.97	9.32	9.38	88.7
10	7.66	9.15	9.38	89.9	86.5
12.5	10.15	9.41	9.50	84.3	83.0
15	12.65	9.56	9.61	81.5	80.8
20	17.64	9.76	9.74	78.1	78.5

[a] The second-order values are those computed from the King-Middleton formula (23) of reference 3; the third-order values are computed from (15).

The second-order functions which occur in (14) are defined as follows:

$$Fc_2 = Fc_2(h) = -\int_0^h K_c(h,z')Fc_{1z'}dz', \quad (18a)$$

$$Fs_2 = Fs_2(h) = -\int_0^h K_s(h,z')Fc_{1z'}dz', \quad (18b)$$

$$Fcs_2 = Fcs_2(h) = -\int_0^h K_c(h,z')Fs_{1z'}dz', \quad (18c)$$

$$Fss_2 = Fss_2(h) = -\int_0^h K_s(h,z')Fs_{1z'}dz'. \quad (18d)$$

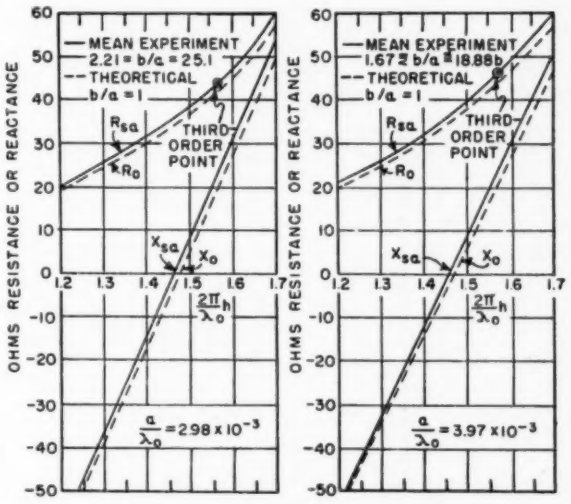


FIG. 1. Resistance and reactance of thin cylindrical antennas near resonance. The theoretical curve is calculated from the second-order formula of King and Middleton, the third-order points are obtained using formula 15 of this paper. The experimental data are those of Hartig; individual points for the several values of b/a differed insignificantly from the mean curve shown.

The expansion parameter is defined at $r=0$ for $\beta_0 h = \pi/2$. It is

$$W = W(0) = \int_0^h L_c(0,z')\cos\beta_0 z'dz'$$

$$= \text{R.P.}[C_a(h,0) - C_a(h,h)]. \quad (19)$$

The real and imaginary parts of the functions $C_a(h,z)$, $S_a(h,z)$, and $E_a(h,z)$ for $\beta_0 h = \pi/2$ are represented, respectively, in Figs. 4, 5; 22, 23; and 24 of reference 3. It is evident from (14) that an accurate determination of the magnitude of $I_p(z)$, when $\beta_0 h = n\pi/2$, n odd, requires at least third-order terms in the denominator. Since the relative distribution of current is not affected by the denominator, consideration of this problem is appropriately reserved for later study in conjunction with the input current and admittance.

For some purposes, for example, the determination of the expansion parameter or electromagnetic fields, an analytically simple approximate representation of the current in a cylindrical antenna is needed. Evidently, the leading terms in the two components of current are given by the following:

$$\frac{I_r(z)}{I_r(0)} = \frac{I_z''}{I_0''} = \left\{\frac{[Fs_1\sin\beta_0 h - Gs_1\cos\beta_0 h]\cos\beta_0 z - [Fs_1(z)\sin\beta_0 h - Gs_1(z)\cos\beta_0 h]\cos\beta_0 h}{[Fs_1\sin\beta_0 h - Gs_1\cos\beta_0 h] - [Fs_1(0)\sin\beta_0 h - Gs_1(0)\cos\beta_0 h]\cos\beta_0 h}\right\}, \quad (20)$$

$$\frac{I_t(z)}{I_t(h-\lambda_0/4)} = \frac{I_z'}{I_m'} = \sin\beta_0(h-z). \quad (21)$$

Except when the antenna involved is long, (20) is unnecessarily complicated and the simpler shifted cosine

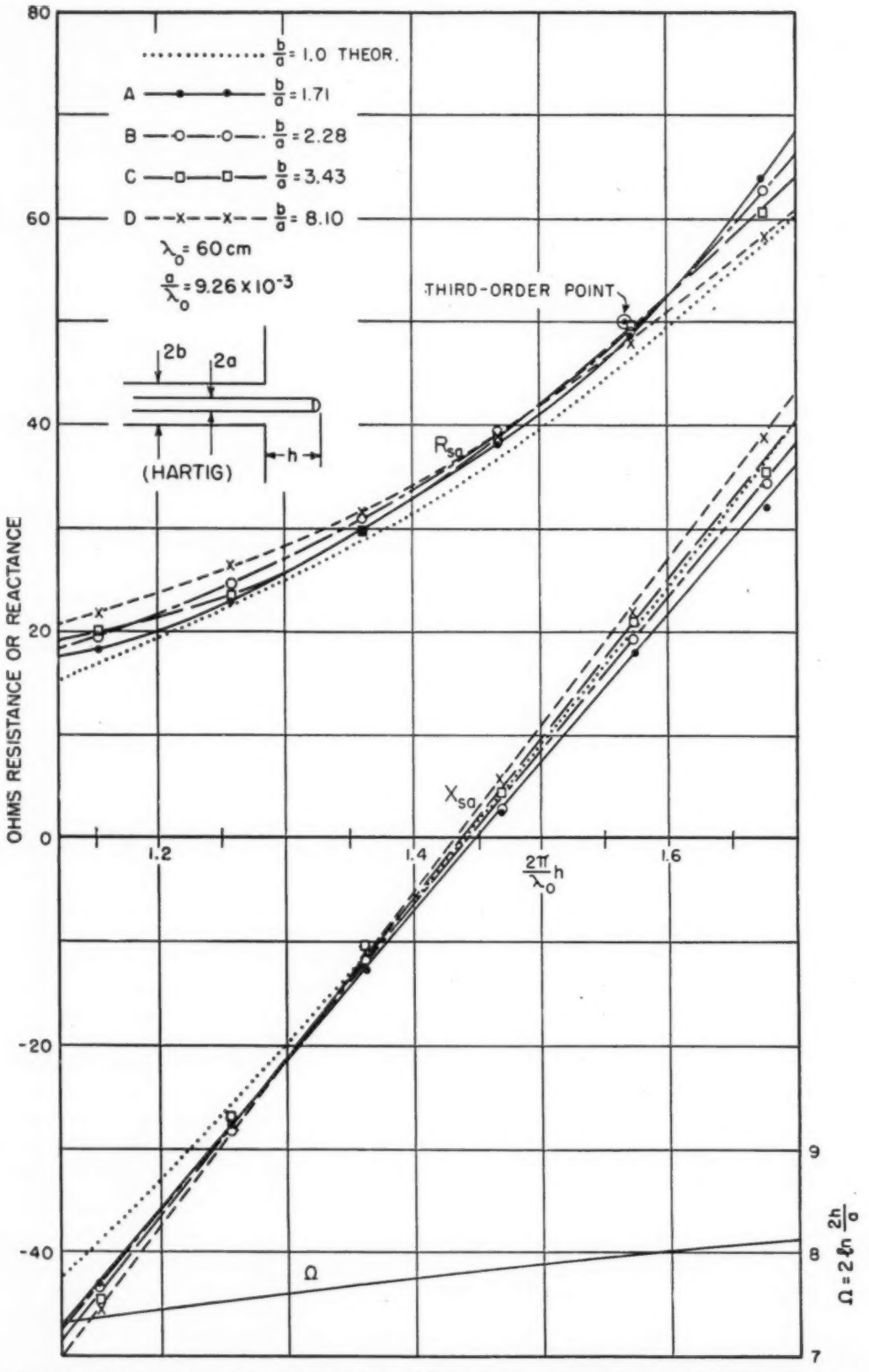


FIG. 2. Resistance and reactance of moderately thick antenna near resonance. The theoretical curve is calculated from the second-order formula of King and Middleton; the third-order point is obtained using formula 15 of this paper. The experimental data are those of Hartig. Note that for the thicker antenna the impedance is quite sensitive to the ratio of b/a.

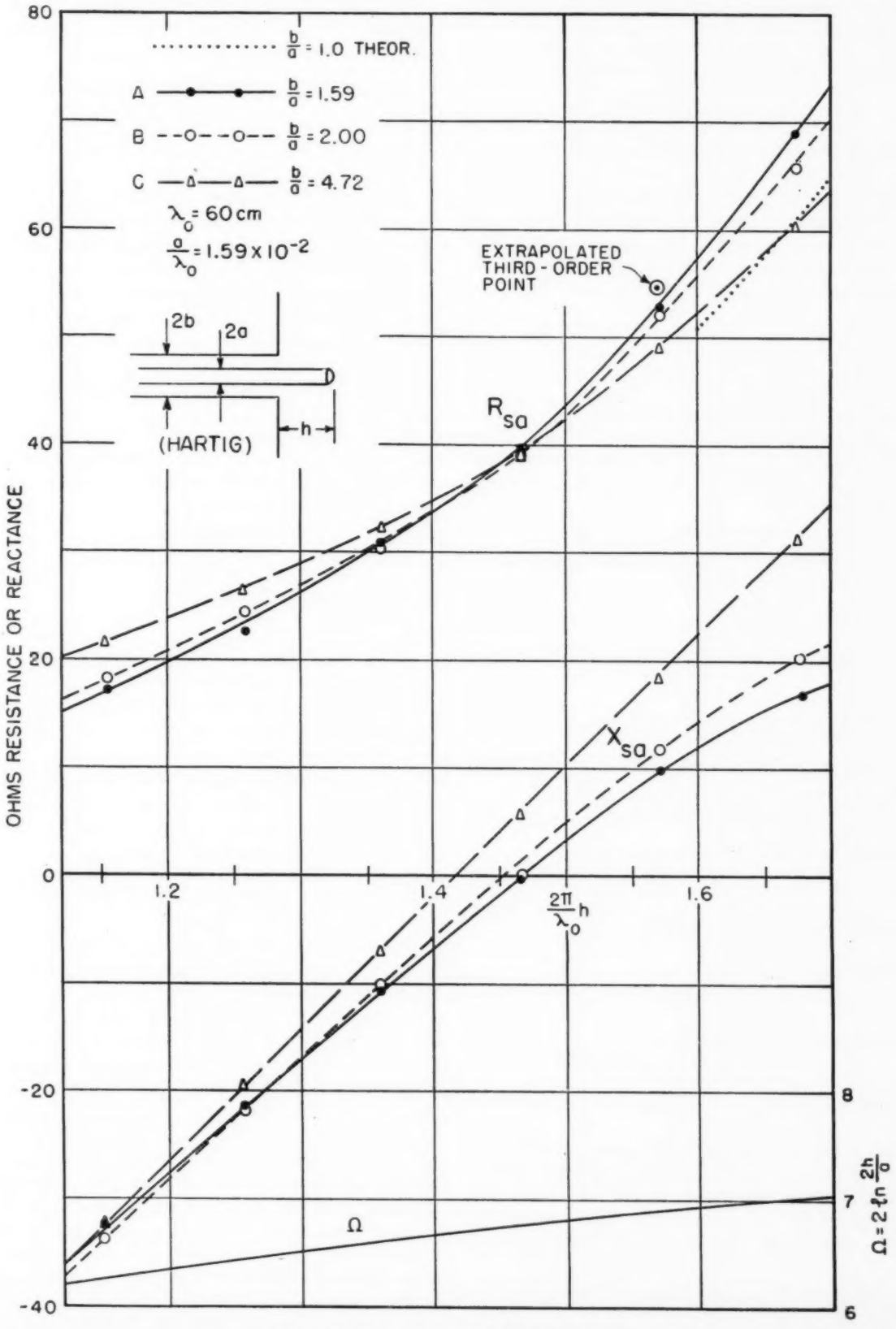


FIG. 3. Like Fig. 2 but for a still thicker antenna.

used as a zeroth-order current in (9a) is adequate. That is, for $\beta_0 h<2\pi$,

$$\frac{I_r(z)}{I_r(0)}=\frac{I_z''}{I_0''}=\frac{\cos\beta_0 z-\cos\beta_0 h}{1-\cos\beta_0 h}. \tag{22}$$

It may be concluded that a relatively simple approximation of the distribution of current in a cylindrical antenna of small radius with $\beta_0 h<2\pi$ is given by $I_z=I_z''+jI_z'$ using (21) and (22). If desired, the ratio of I_0'' to I_m' may be obtained from first-order data. When $\beta_0 h\leqq\pi/2$, the distributions (21) and (22) are sufficiently alike so that the single formula

$$I_z=I_0\frac{\sin\beta_0(h-z)}{\sin\beta_0 h} \tag{23}$$

is satisfactory for the *complex* current.

Throughout this section currents in the range $h\leqq z\leqq 0$ have been determined. Currents in the range $-h\leqq z\leqq 0$ are obtained directly using the symmetry condition $I(-z)=I(z)$.

Distributions of charge corresponding to the simple currents (21) and (22) may be obtained using the equation of continuity, $dI/dz=j\omega q=0$.

CONDUCTANCE AND RESISTANCE FOR $h=\lambda_0/4$

From separate integral equations for the components of current in phase and in phase-quadrature with the driving voltage, it is shown in the preceding that the leading term in the denominator of the formula (13a) for the in-phase component I_z'' is of second order when $\beta_0 h$ is at or very near $n\pi/2$, n odd, since zeroth- and first-order terms vanish. It may be anticipated, therefore, that quantities which depend primarily or entirely on the in-phase current may be determined less accurately by a second-order fromula for these values of $\beta_0 h$ than for others. Evidently, this is true particularly of the conductance G_0 which has a sharp maximum near $\beta_0 h=n\pi/2$, n odd.

It is possible to investigate the conductance at $\beta_0 h=n\pi/2$, n odd, using terms up to and including those of the third-order in the small denominator by setting $z=0$ in (14). The resulting expressing is

$$G_0=\frac{2\pi}{\zeta_0}\left\{\frac{[Fs_1+W^{-1}(Fs_1Gc_1-Gs_1Fc_1)+W^{-1}[Fc_1(0)Fs_1-Fs_1(0)Fc_1]}{[Fc_1^2+Fs_1^2]+2W^{-1}[Fc_1(Fc_2-Fss_2)+Fs_1(Fs_2+Fcs_2)]}\right\}, \tag{15}$$

where, for $\beta_0 h=\pi/2$, and using the notation of reference 3,

$$Gs_1=Fc_1=\alpha_1{}^I=\tfrac{1}{2}\,\mathrm{Si}2\pi=-0.709 \tag{16a}$$

$$Fs_1=\alpha_1{}^{II}=\tfrac{1}{2}\,\mathrm{Cin}2\pi=1.219, \tag{16b}$$

$$Gc_1=\alpha_1{}^{II}+2\ln 2=0.1675, \tag{16c}$$

$$Fc_1(0)=W-\mathrm{R.P.}C_a(h,0)=-\mathrm{R.P.}C_a(h,h)=\alpha_1{}^I=-0.709, \tag{16d}$$

$$Fs_1(0)=-\mathrm{I.P.}C_a(h,0)=1.852, \tag{16e}$$

$$(Fc_2-Fss_2)=\alpha_2{}^I+(W-\Omega)\alpha_1{}^I=-3.19+0.709(\Omega-W), \tag{17a}$$

$$(Fs_2+Fcs_2)=\alpha_2{}^{II}+(W-\Omega)\alpha_1{}^{II}=2.65-1.219(\Omega-W), \tag{17b}$$

$$W=W(0)=C_a(h,0)-C_a(h,h). \tag{18}$$

Numerical values of the expansion parameter W for a series of values of $\Omega=2\ln(2h/a)$ are in Table I. Third-order† values of G_0 computed from (15) also are in this table together with the second-order values calculated from (23) of reference 3. Since the second-order reactance is known from reference 1 to be quite accurate at $\beta_0 h=\pi/2$, the values of resistance corresponding to the newly determined third-order values of G_0 also may be evaluated. These are in Table I together with the previously determined second-order values. It is seen that the new and more accurate formula (15) gives values of the conductance at $\beta_0 h=\pi/2$ that are somewhat lower, and of the resistance that are somewhat higher than those obtained previously, the difference being greater for thicker antennas and relatively unimportant for $\Omega>10$; at $\Omega=10$ the difference is about 4 percent. It is evident from this analysis that the theoretical conductances computed from the second-order formula (23) of reference 3 are somewhat too high in the range of resonance.

Although only the one point at $\beta_0 h=\pi/2$ is evaluated readily from the third-order formula, it is interesting to compare this point with the previously evaluated complete second-order data and the extensive experimental results of Hartig.[5] These are obtained on an antenna of length h and radius a that is the extension above a 36×36-foot ground screen of the inner conductor of a coaxial line with an outer conductor of inner radius b. Impedance curves as a function of $\beta_0 h$ with a/λ_0 as parameter are shown in Figs. 1 through 3 for a range of values of b/a. The theoretical curves correspond to the ideal limiting experimental curves for $b/a=1$. While

† The numerator in (15) actually includes only zeroth- and first-order terms. However, since the contribution of the first-order terms is only 0.8 percent for $\Omega=7$ and 0.2 percent for $\Omega=20$, it is evident that contributions from second- and third-order terms must be negligible so that the results are, in effect, of third order.

[5] E. O. Hartig, "Circular Apertures and their Effects on Half-Dipole Impedances," Cruft Laboratory Technical Report No. 107 (June, 1950).

the second-order resistances compare quite well with experiment in the range near resonance, especially for the thinner antennas, the new third-order points at $\beta_0 h = \pi/2$ are in excellent agreement even for the thicker antennas.

It may be concluded that by improving the accuracy in the determination of the component of current in phase with the driving voltage, accurate values of conductance and resistance near resonance may be obtained with a third-order iteration.

JOURNAL OF APPLIED PHYSICS VOLUME 24, NUMBER 2 FEBRUARY, 1953

The Motion of a Sphere Moving Parallel to a Plane Boundary*

F. C. KARAL†
Defense Research Laboratory, The University of Texas, Austin, Texas
(Received August 15, 1952)

The motion of a rigid sphere moving through an incompressible, homogeneous, nonviscous fluid in the presence of a rigid, infinite plane boundary is examined. The direction of motion of the sphere is assumed to be parallel to the plane boundary. It is assumed that the motion is irrotational and that the sphere velocity is small. An exact treatment is presented. Several curves of the pressure variation on the plane boundary are plotted for different values of the ratio of sphere radius to the perpendicular distance from the center of the sphere to the plane boundary.

I. INTRODUCTION

THE motion of a rigid sphere moving through an incompressible, homogeneous, nonviscous fluid in the presence of a rigid, infinite plane boundary is examined. The direction of motion of the sphere is assumed to be parallel to the plane boundary. The fluid is bounded below by the plane boundary but is assumed to be of infinite extent above the plane boundary. It is assumed that the motion is irrotational and that the sphere velocity is small.

Lamb[1] gives a solution to the problem; however, his results are based upon the assumption that the ratio of sphere radius to the perpendicular distance from the center of the sphere to the plane boundary is small. A search of the literature for an exact solution has been unsuccessful. The treatment given here is exact and is believed to be new.

Consider a stationary, rigid sphere located a finite distance from a rigid, infinite plane boundary and immersed in a fluid moving in the negative x direction with velocity V. The problem is to find an expression for the steady state pressure on the plane boundary as a function of the dimensionless parameter x/h, where x is the distance measured along the positive x axis from a point directly beneath the center of the sphere, and h is the perpendicular distance from the center of the sphere to the plane boundary. We are actually interested in the problem in which the sphere moves with constant velocity V in the x direction, and the fluid, at infinity, is at rest. The pressure distribution is the same in either case, so the stationary sphere representation is preferred since the analytical details are simpler. In the sequel we shall deal with the stationary sphere problem.

II. THE COORDINATE SYSTEM

The velocity potential ϕ satisfies Laplace's equation

$$\nabla^2\phi = 0. \tag{1}$$

Solutions of (1) in rotated bipolar coordinates (ξ, η, φ) are given by

$$\begin{aligned}\phi(\xi, \eta, \varphi) = \{\cosh\eta - \cos\xi\}^{\frac{1}{2}}\{&AP_n^m(\cos\xi) \\ &+ BQ_n^m(\cos\xi)\}\{C\sinh(n+\tfrac{1}{2})\eta \\ &+ D\cosh(n+\tfrac{1}{2})\eta\}\{E\sin m\varphi + F\cos m\varphi\},\end{aligned} \tag{2}$$

where $P_n^m(\cos\xi)$ and $Q_n^m(\cos\xi)$ are the associated Legendre polynomials of the first and second kind, respectively.[2] The transformations to rectangular coordinates are given by

$$\left.\begin{aligned} x &= r\cos\varphi \\ y &= r\sin\varphi \\ z &= \frac{a\sinh\eta}{\cosh\eta - \cos\xi}\end{aligned}\right\}, \tag{3}$$

where

$$r = \frac{a\sin\xi}{\cosh\eta - \cos\xi}. \tag{4}$$

The coordinate system is illustrated in Fig. 1. Note that the rotated bipolar coordinate system differs from the toroidal coordinate system in that we obtain spheres, tores (without holes) and planes as the orthogonal

* Research done under Bureau of Ordnance Research and Development Contract NOrd-10639.
† Now at Magnolia Petroleum Company, Field Research Laboratories, Dallas, Texas.
[1] H. Lamb, *Hydrodynamics* (Dover Publications, New York, 1945), sixth edition, pp. 133, 134.
[2] E. W. Hobson, *Theory of Spherical and Ellipsoidal Harmonics* (Cambridge University Press, Cambridge, 1931), Chapter III.

coordinate surfaces instead of intersecting spheres, tores (with holes) and planes.

The geometry of the problem can be fitted to a rotated bipolar coordinate system in a natural and easy way. The sphere will be denoted by the coordinate surface $\eta=\eta_0$; the plane boundary will be denoted by the coordinate surface $\eta=0$ (see Fig. 2).

III. THE BOUNDARY CONDITIONS

The potential ϕ, in addition to being a solution of Laplace's equation, must satisfy the following boundary conditions:

$$\phi\rightarrow+Vr\cos\varphi, \quad r\rightarrow\infty, \tag{5}$$

$$v_\eta=0, \quad \eta=0, \tag{6}$$

$$v_\eta=0, \quad \eta=\eta_0. \tag{7}$$

The particle velocity $\mathbf{v}$ can be found from the relation

$$\mathbf{v}=-\nabla\phi. \tag{8}$$

In order to satisfy the boundary conditions (5)–(7), we write

$$\phi(\xi,\eta,\varphi)=\frac{Va\sin\xi}{\cosh\eta-\cos\xi}\cos\varphi+\sum_{n=1}^{\infty}A_n(\cosh\eta-\cos\xi)^{\frac{1}{2}}P_n^1(\cos\xi)\times\cosh(n+\tfrac{1}{2})\eta\cos\varphi. \tag{9}$$

Boundary conditions (5) and (6) can readily be shown to be satisfied; boundary condition (7) can be satisfied by suitable choice of the constants A_n. Since

$$v_\eta=-\frac{\cosh\eta-\cos\xi}{a}\frac{\partial\phi}{\partial\eta}, \tag{10}$$

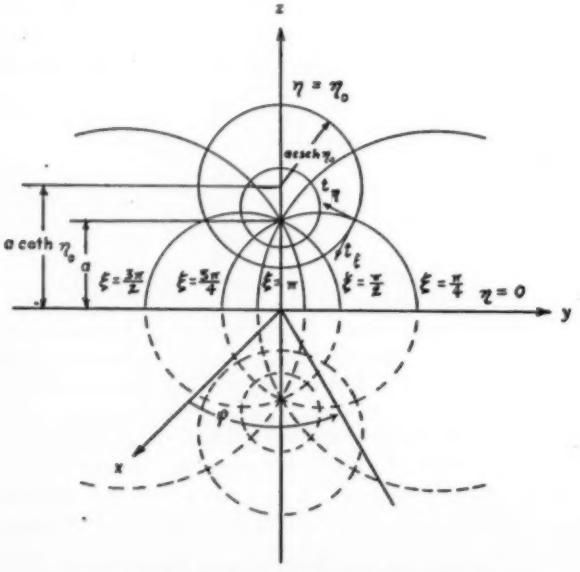


FIG. 1. The rotated bipolar coordinate system.

boundary condition (7) becomes

$$-\frac{Va\sin\xi\sinh\eta_0}{(\cosh\eta_0-\cos\xi)^2}\cos\varphi=+\sum_{n=1}^{\infty}(n+\tfrac{1}{2})A_n(\cosh\eta_0-\cos\xi)^{\frac{1}{2}}P_n^1(\cos\xi)\times\sinh(n+\tfrac{1}{2})\eta_0\cos\varphi+\sum_{n=1}^{\infty}\tfrac{1}{2}A_n\frac{\sinh\eta_0}{(\cosh\eta_0-\cos\xi)^{\frac{1}{2}}}P_n^1(\cos\xi)\times\cosh(n+\tfrac{1}{2})\eta_0\cos\varphi. \tag{11}$$

Making use of the well-known relation[3]

$$\cos\xi P_n^1(\cos\xi)=\frac{(n+1)}{(2n+1)}P_{n-1}^1(\cos\xi)+nP_{n+1}^1(\cos\xi), \tag{12}$$

we find that Eq. (11) can be written in the form

$$\frac{2Va\sin\xi\sinh\eta_0}{(\cosh\eta_0-\cos\xi)^{\frac{1}{2}}}=\sum_{n=1}^{\infty}P_n^1(\cos\xi)\left[A_n\left\{2n\cosh\eta_0\sinh\frac{2n+1}{2}\eta_0+\sinh\frac{2n+3}{2}\eta_0\right\}-A_{n+1}(n+2)\sinh\frac{2n+3}{2}\eta_0-A_{n-1}(2n-1)(n-1)\sinh\frac{2n-1}{2}\eta_0\right]. \tag{13}$$

In order to eliminate the variable ξ it is necessary to perform the following expansion:

$$\frac{\sin\xi}{(\cosh\eta_0-\cos\xi)^{\frac{1}{2}}}=\sum_{n=1}^{\infty}a_nP_n^1(\cos\xi). \tag{14}$$

This evaluation is carried out in the Appendix. Substituting (14) into (13), we have

$$2Va\sinh\eta_0\sum_{n=1}^{\infty}a_nP_n^1(\cos\xi)=\sum_{n=1}^{\infty}P_n^1(\cos\xi)\left[A_n\left\{2n\cosh\eta_0\sinh\frac{2n+1}{2}\eta_0+\sinh\frac{2n+3}{2}\eta_0\right\}-A_{n+1}(n+2)\sinh\frac{2n+3}{2}\eta_0-A_{n-1}(2n-1)(n-1)\sinh\frac{2n-1}{2}\eta_0\right]. \tag{15}$$

[3] See reference 2, p. 108.

Equating like coefficients of $P_n^1(\cos\xi)$, we obtain

$$(2Va)a_n \sinh\eta_0 = A_n\left\{2n\cosh\eta_0 \sinh\frac{2n+1}{2}\eta_0 + \sinh\frac{2n+3}{2}\eta_0\right\} - A_{n+1}(n+2)\sinh\frac{2n+3}{2}\eta_0 - A_{n-1}(2n-1)(n-1)\sinh\frac{2n-1}{2}\eta_0, \quad (16)$$

where $n \geq 1$.

Equation (16) consists of an infinite number of equations, since n ranges from 1 to ∞. These equations can be solved by approximation methods to obtain a finite number of the constants A_n to any specified degree of accuracy.

IV. THE PRESSURE EQUATION

When a velocity potential exists and the motion is steady, the equations of motion can be integrated to yield[4]

$$\frac{p}{d} = -\Omega - \tfrac{1}{2}v^2 + C, \quad (17)$$

where Ω is the gravitational potential, C is a constant, d is the density, and v is the velocity of the fluid. Now

$$v^2 = v_\xi^2 + v_\eta^2 + v_\varphi^2 \quad (18)$$

or

$$v^2 = \frac{(\cosh\eta - \cos\xi)^2}{a^2} \times\left[\left(\frac{\partial\phi}{\partial\xi}\right)^2 + \left(\frac{\partial\phi}{\partial\eta}\right)^2 + \frac{1}{\sin^2\xi}\left(\frac{\partial\phi}{\partial\varphi}\right)^2\right]. \quad (19)$$

We shall determine the pressure along the positive x axis and on the surface of the rigid plane boundary. Under these conditions, $\eta=0$ and $\varphi=0$. Equation (19) simplifies to

$$v^2 = \frac{(1-\cos\xi)^2}{a^2}\left(\frac{\partial\phi}{\partial\xi}\right)^2. \quad (20)$$

From Eq. (9) we obtain

$$\frac{\partial\phi}{\partial\xi} = \frac{Va\cos\xi}{(\cosh\eta-\cos\xi)}\cos\varphi - \frac{Va\sin^2\xi}{(\cosh\eta-\cos\xi)^2}\cos\varphi + \sum_{n=1}^{\infty}\tfrac{1}{2}A_n\frac{\sin\xi}{(\cosh\eta-\cos\xi)^{\frac{1}{2}}}P_n^1(\cos\xi) \times\cosh(n+\tfrac{1}{2})\eta\cos\varphi + \sum_{n=1}^{\infty}A_n(\cosh\eta-\cos\xi)^{\frac{1}{2}} \times\frac{\partial P_n^1(\cos\xi)}{\partial\xi}\cosh(n+\tfrac{1}{2})\eta\cos\varphi. \quad (21)$$

[4] See reference 1, p. 21.

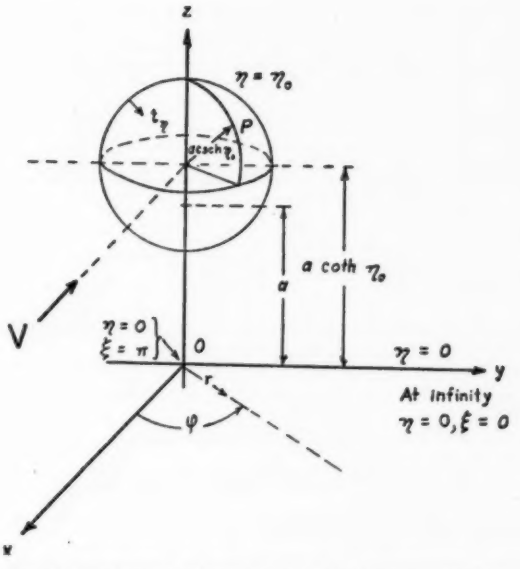


FIG. 2. Stationary sphere, moving fluid representation.

Performing the indicated differentiation, setting $\eta=\varphi=0$, and collecting terms, we have

$$\frac{\partial\phi}{\partial\xi} = -\frac{Va}{(1-\cos\xi)} + \frac{\sin\xi}{2(1-\cos\xi)^{\frac{1}{2}}}\sum_{n=1}^{\infty}A_nP_n^1(\cos\xi) + \frac{(1-\cos\xi)^{\frac{1}{2}}}{\sin\xi}\sum_{n=1}^{\infty}(n)A_nP_{n+1}^1(\cos\xi) - \cot\xi(1-\cos\xi)^{\frac{1}{2}}\sum_{n=1}^{\infty}(n+1)A_nP_n^1(\cos\xi). \quad (22)$$

For the purposes of this problem, it is found that sufficient accuracy is obtained by considering A_1 and A_2 only. Equation (22) becomes

$$\frac{\partial\phi}{\partial\xi} = -\frac{Va}{(1-\cos\xi)} + \frac{\sin\xi}{2(1-\cos\xi)^{\frac{1}{2}}}[A_1P_1^1(\cos\xi)+A_2P_2^1(\cos\xi)] + (1-\cos\xi)^{\frac{1}{2}}\left[A_1\frac{P_2^1(\cos\xi)}{\sin\xi} + A_2\frac{2P_3^1(\cos\xi)}{\sin\xi}\right] - \cos\xi(1-\cos\xi)^{\frac{1}{2}}\left[2A_1\frac{P_1^1(\cos\xi)}{\sin\xi} + 3A_2\frac{P_2^1(\cos\xi)}{\sin\xi}\right]. \quad (23)$$

Substituting for the associated Legendre polynomials, and rearranging terms, we obtain

$$\frac{\partial\phi}{\partial\xi} = -\frac{Va}{(1-\cos\xi)} + \frac{1}{4(1-\cos\xi)^{\frac{1}{2}}} \times\{A_1F_1(\xi)+3A_2F_2(\xi)\}, \quad (24)$$

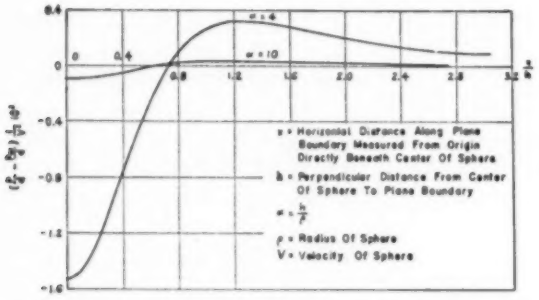

x = Horizontal Distance Along Plane Boundary Measured From Origin Directly Beneath Center Of Sphere
h = Perpendicular Distance From Center Of Sphere To Plane Boundary
ρ = Radius Of Sphere
V = Velocity Of Sphere

FIG. 3. Dimensionless pressure on a plane boundary as a function of the dimensionless parameter x/h.

where

$$F_1(\xi)=-3\cos 2\xi+4\cos\xi-1, \tag{25}$$

$$F_2(\xi)=+4\cos 2\xi+\cos\xi-5\cos\xi\cos 2\xi. \tag{26}$$

Equation (20) now becomes

$$v^2=V^2-\frac{(1-\cos\xi)^{\frac{1}{2}}}{2}V^2\left[\left(\frac{A_1}{Va}\right)F_1(\xi)+3\left(\frac{A_2}{Va}\right)F_2(\xi)\right]$$

$$+\frac{(1-\cos\xi)}{16}V^2\left[\left(\frac{A_1}{Va}\right)F_1(\xi)+3\left(\frac{A_2}{Va}\right)F_2(\xi)\right]^2. \tag{27}$$

Substituting the above expression into the equation of motion (17), we have

$$\left(\frac{p}{d}-\frac{p_\infty}{d}\right)\frac{1}{V^2}=L(\xi)-\tfrac{1}{2}[L(\xi)]^2, \tag{28}$$

where

$$L(\xi)=\frac{(1-\cos\xi)^{\frac{1}{2}}}{4}\left[\left(\frac{A_1}{Va}\right)F_1(\xi)+3\left(\frac{A_2}{Va}\right)F_2(\xi)\right]. \tag{29}$$

We have also set the gravitational potential equal to zero and evaluated the constant C by assuming that the pressure at infinity is equal to $\frac{1}{2}dV^2$.

V. CALCULATIONS

We wish to plot the pressure on the surface of the rigid plane boundary as a function of the dimensionless parameter x/h. The pressure will be found for two values of the ratio h/ρ, where ρ is the radius of the sphere. We shall let $\alpha=h/\rho=\cosh\eta_0$ for convenience.

The method of calculation is straightforward but tedious. For selected values of α, it is first necessary to determine the constants a_n. Knowing these constants, the coefficients A_n can be found. Once the A_n's are known, the pressure can be computed from Eq. (28). The pressure can be plotted as a function of x/h by using the relation

$$x/h=\tanh\eta_0\cot\xi/2. \tag{30}$$

For purposes of calculation, values of α equal to 4 and 10 have been selected. It is found for these values of α that the square of $L(\xi)$ can be neglected in (28). Therefore

$$(p-p_\infty)1/V^2=L(\xi), \tag{31}$$

where V is measured in cm/sec, d is measured in g/cm³, and the pressure is expressed in dynes/cm². In practical units the pressure equation for water can be written as

$$(p-p_\infty)=0.382V^2L(\xi), \tag{32}$$

where V is measured in ft/sec and the pressure is expressed in inches of water. From (29) we find, for the case when $\alpha=4$, that

$$L(\xi)=1.48(10)^{-3}(1-\cos\xi)^{\frac{1}{2}}$$
$$\times[F_1(\xi)+8.32(10)^{-2}F_2(\xi)], \tag{33}$$

$$x/h=0.968\cot\xi/2. \tag{34}$$

From (29) we find, for the case when $\alpha=10$, that

$$L(\xi)=8.90(10)^{-5}(1-\cos\xi)^{\frac{1}{2}}$$
$$\times[F_1(\xi)+1.29(10)^{-2}F_2(\xi)], \tag{35}$$

$$x/h=0.955\cot\xi/2. \tag{36}$$

Figure 3 is a plot of the dimensionless quantity $(p-p_\infty)(1/dV^2)$ as a function of x/h for α equal to 4 and 10. Figures 4 and 5 are plots of the pressure as a function of x/h for V equal to 16, 20, and 24 ft/sec, and α equal to 4 and 10, respectively. These curves are valid whether the sphere is at rest and the fluid streams by, or the sphere moves forward in steady motion and the fluid at infinity is at rest. It should be noted that the problem just solved is equivalent to the problem of two spheres moving in an incompressible, homogeneous, nonviscous fluid with the same velocity V perpendicular to their line of centers. Equation (31) gives the pressure along the perpendicular bisector of the line of centers and in the direction of motion of the spheres.

It is interesting to compare the approximate solution given by Lamb with the exact solution given here. At the point of maximum negative pressure, directly beneath the sphere, the percentage error between the two results is roughly 2 percent and $\frac{1}{2}$ percent for α equal to 4 and 10, respectively. The approximate solution yields results that are too large.

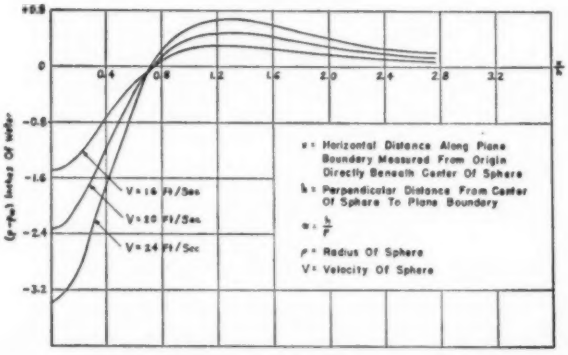

x = Horizontal Distance Along Plane Boundary Measured From Origin Directly Beneath Center Of Sphere
h = Perpendicular Distance From Center Of Sphere To Plane Boundary
ρ = Radius Of Sphere
V = Velocity Of Sphere
V = 16 Ft/Sec
V = 20 Ft/Sec
V = 24 Ft/Sec

FIG. 4. Pressure on a plane boundary as a function of the dimensionless parameter x/h for $\alpha=4$.

VI. ACKNOWLEDGMENT

The writer wishes to thank Mr. Wilmer Hoyer for help in computing the curves and Dr. C. W. Horton for offering several suggestions.

APPENDIX

The coefficients a_n appearing in Eq. (14) will now be evaluated. It is assumed that the left-hand side of (14) can be expanded into an associated Legendre series. By using the orthogonality formulas

$$\int_{-1}^{+1}[P_n^m(\zeta)]^2 d\zeta=\frac{2}{(2n+1)}\frac{(n+m)!}{(n-m)!}, \tag{1}$$

$$\int_{-1}^{+1}P_n^m(\zeta)P_p^m(\zeta)d\zeta=0, \quad n\neq p, \tag{2}$$

we obtain

$$a_n=\frac{(2n+1)}{2}\frac{(n-1)!}{(n+1)!}I_n(\alpha), \tag{3}$$

where

$$I_n(\alpha)=\int_{-1}^{+1}\frac{(1-\zeta^2)^{\frac{1}{2}}}{(\alpha-\zeta)^{\frac{1}{2}}}P_n^1(\zeta)d\zeta, \tag{4}$$

$$\alpha=\cosh\eta, \tag{5}$$

$$\zeta=\cos\xi. \tag{6}$$

From the definitions of the associated Legendre polynomials, it is evident that we can express (4) in terms of integrals of the form

$$F_k(\alpha)=\int_{-1}^{+1}\frac{\zeta^k}{(\alpha-\zeta)^{\frac{1}{2}}}d\zeta. \tag{7}$$

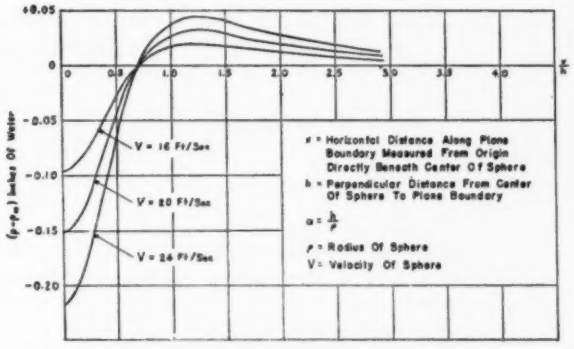


FIG. 5. Pressure on a plane boundary as a function of the dimensionless parameter x/h for $\alpha=10$.

Let

$$F_k(\alpha)=G_k(\alpha,+1)-G_k(\alpha,-1), \tag{8}$$

where

$$G_k(\alpha,\zeta)=\int\frac{\zeta^k}{(\alpha-\zeta)^{\frac{1}{2}}}d\zeta, \quad k=1,2,3\cdots. \tag{9}$$

Consider

$$L_k(\alpha,\zeta)=\int\frac{\zeta^k}{(\alpha-\zeta)^{\frac{1}{2}}}d\zeta. \tag{10}$$

Integrating (9) and (10) by parts we obtain

$$G_k(\alpha,\zeta)=\frac{2\zeta^k}{(\alpha-\zeta)^{\frac{1}{2}}}-2kL_{k-1}(\alpha,\zeta) \tag{11}$$

and

$$L_k(\alpha,\zeta)=-\frac{2\zeta^k}{(2k+1)}(\alpha-\zeta)^{\frac{1}{2}}+\frac{2k\alpha}{(2k+1)}L_{k-1}(\alpha,\zeta). \tag{12}$$

Since

$$L_0(\alpha,\zeta)=-2(\alpha-\zeta)^{\frac{1}{2}}, \tag{13}$$

we can find $L_k(\alpha,\zeta)$ for any k from (12). $I_n(\alpha)$ can be expressed in terms of $L_k(\alpha)$ by using (8) and (11).

The Effect of the Rate of Cooling on the Allotropic Transformation Temperatures of Uranium*

POL DUWEZ
California Institute of Technology, Pasadena, California, and Atomic Energy Research Department, North American Aviation, Inc., Downey, California

(Received August 11, 1952)

Allotropic transformation in uranium has been studied over a range of cooling rates from 5 to about 8000°C/sec. The transformation temperatures of both gamma-to-beta and beta-to-alpha were found to decrease continuously with increasing rates of cooling. The extent of the beta-range increased with increasing cooling rates. For rates of cooling up to 1000°C/sec, recalescence was observed in both transformations. For higher cooling rates, there was usually no recalescence.

In most of the recorded cooling curves, a small but definite thermal arrest was observed, between the two main arrests which correspond to the two known phase transformations. This additional thermal arrest was also present in a heating curve, where it occurred at about 740°C, compared with 666 and 771°C for the two known phase transformations. Possible explanations of the additional arrest are discussed.

INTRODUCTION

URANIUM undergoes allotropic transformations in the vicinity of 665 and 770°C. The low temperature form, alpha, has an orthorhombic symmetry[1] and the high temperature form, gamma, is body-centered cubic.[2] The crystal structure of the beta-form (stable between 665 and 770°C) has been recently found to be tetragonal, with 30 atoms per unit cell.[3] This structure is apparently isomorphous with that of the sigma-phase[4] found in several binary alloy systems involving the transition elements.[5–7]

TABLE I. Data on allotropic transformation of uranium.

Investigator	Transformation temperature (°C) gamma-to-beta	beta-to-alpha
Chipman (reference 8)	775	665
Dahl-Van Dusen (reference 9)		
cooling	764	645
heating	772	667
Moore-Kelley (reference 10)	772	662
Gordon-Kaufmann (reference 11)	765	772
Pfeil (reference 12)	772	665
Grogan (reference 13)		
cooling	770	660
heating	780	675
	Latent heat of transformation (cal/g at)	
Moore-Kelley (reference 10)	1165	680

The temperatures at which the alpha-to-beta and the beta-to-gamma transformations take place have been determined by several investigators.[8–13] Probably because of differences in the purity of the material as well as in the cooling or heating rate used by the various experimenters, the measured temperatures agree within only about ±10°C (see Table I). From the results published so far, however, it is not possible to ascertain whether or not the transition temperatures can be substantially depressed by increasing the rate of cooling, although it is well known that neither of the high temperature phases can be retained by quenching the pure metal to room temperature. The present investigation deals with the measurement of the allotropic transformation temperatures with rates of cooling varying from 5 to about 8000°C/sec.

EXPERIMENTAL PROCEDURE

The method for measuring transformation temperature at very fast rates of cooling consists essentially of heating a small sample of metal to which thermocouple wires are spot welded, cooling the sample by means of a jet of helium and recording the temperature *vs* time curve on an oscillograph. The apparatus used for the present experiments was similar to that described by Greninger in his study of the martensite arrest in carbon steels.[14] The retractable heating coil was made of molybdenum wire (0.04 inch in diameter) and the assembly was placed in a bell jar with a vacuum of 10^{-5} mm Hg or better.

* This paper is based on studies conducted for the U. S. Atomic Energy Commission on Contract No. AT-11-1-GEN-8. The experiments were performed with equipment constructed at California Institute of Technology under Contract No. N6onr-24430 with the Office of Naval Research, U. S. Navy.

[1] C. W. Jacob and B. E. Warren, J. Am. Chem. Soc. **59**, 2588–2591 (1937).
[2] A. S. Wilson and R. E. Rundle, Acta Cryst. **2**, 148–150 (1949).
[3] C. W. Tucker, Jr., Science **112**, 448 (1950).
[4] Dickins, Douglas, and Taylor, J. Iron and Steel Inst. **167**, 27 (1951).
[5] D. P. Shoemaker and B. G. Bergman, J. Am. Chem. Soc. **72**, 5793 (1950).
[6] P. Duwez and S. R. Baen, "X-Ray Study of the Sigma Phase in Various Alloy Systems," *Symposium on the Nature, Occurrence, and Effects of the Sigma Phase*, Spec. Tech. Pub. No. 110, Am. Soc. Testing Materials, Philadelphia, pp. 48–60, 1951.
[7] Pearson, Christian, and Hume-Rothery, Nature **167**, 110 (1951).
[8] J. Chipman, "Metallurgy in the Development of Atomic Power," U. S. Atomic Energy Commission Pub. MDDC, (1946).
[9] A. I. Dahl and M. S. Van Dusen, J. Research, Natl. Bur. Standards **39**, 53–58 (1947).
[10] G. E. Moore and K. K. Kelley, J. Am. Chem. Soc. **69**, 2105–2107 (1947).
[11] P. Gordon and A. R. Kaufmann, Trans. Am. Inst. Mining Met. Engrs. **188**, 182–194 (1950).
[12] P. C. L. Pfeil, J. Inst. of Metals **77**, 553–570 (1950).
[13] J. D. Grogan, J. Inst. of Metals **77**, 571–580 (1950).
[14] A. B. Greninger, Trans. Am. Soc. Metals, **30**, 1, 1942.

The specimens were originally made of two pieces of uranium about 0.030 inch thick and less than $\frac{1}{16}$ inch square. The 0.005-inch platinum-platinum rhodium or chromel-alumel thermocouple wires were placed between the two pieces of metal and the assembly spot welded. It was later found that when a high speed helium blast was used the two pieces were not always cooled at the same rate (probably because of the lack of symmetry in gas flow) and extraneous effects were observed. Specimens were then prepared by spot welding one small piece of uranium between two very thin platinum sheets of the same size (about $\frac{1}{16}$ inch square). In this technique, the thermocouple wires were inserted, before spot welding, between the uranium sample and the platinum covers.

The temperature was recorded on a rotating drum oscillograph.† The galvanometer was a 65-cycle-per-second element, having a resistance of 45 ohms and a sensitivity of 1-inch deflection per 25 microamperes. The instrument was calibrated by means of a Leeds and Northrup portable precision potentiometer. Records were obtained with paper speed of about $4\frac{1}{2}$ inches per second. With the exception of a few tests, a thermal arrest in the cooling curve at the transformation temperature was quite easy to locate, and the precision of measurement of the transformation temperature was approximately ± 4°C.

The quality of uranium metal used in this investigation was equivalent to that of "high purity" uranium metal, such as is available, for example, from Mallinckrodt. After most of the experiments were performed, a sample of metal was obtained which was stated to be at least as pure as the previous lot and probably of higher purity. A few additional measurements were then performed on this material, which is designated as "higher purity" metal.

RESULTS OF MEASUREMENTS ON HIGH PURITY URANIUM

About 50 cooling curves were obtained with rates of cooling ranging from 10 to 8000°C/sec. Typical curves

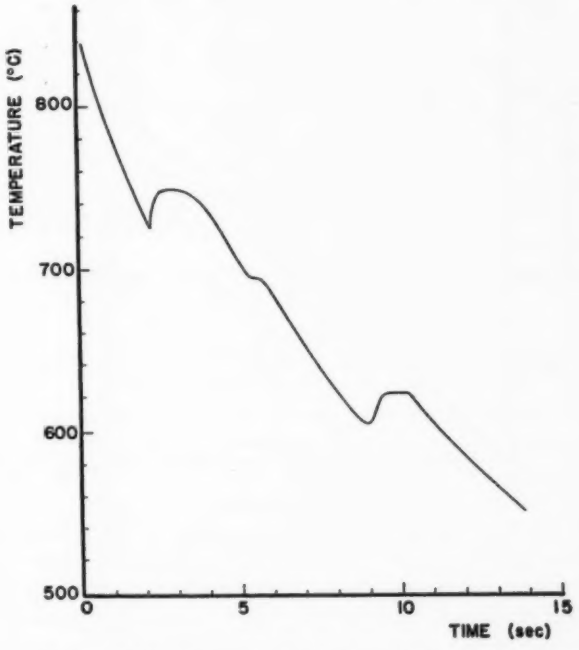


FIG. 1. Cooling curve obtained at a rate of 120°C/sec.

are shown in Figs. 1, 2, and 3. For rates of cooling up to about 1000°C/sec, a very definite recalescence was observed in many experiments, both in gamma-to-beta and in beta-to-alpha transformations (see Figs. 1 and 2). For higher rates of cooling, there was usually no recalescence, but the thermal arrests were clearly defined, as shown in Fig. 3. In the graph of Fig. 4, showing the effect of rate of cooling in the transformation temperature, recalescence is shown by a vertical line connecting the minimum temperatures corresponding to the beginning of transformation to the maximum temperature reached after recalescence. The presence of recalescence is obviously one of the reasons why the measured transformation temperatures, shown on the graph of Fig. 4, exhibit quite a large scatter. However, even at high rates of cooling, for which recalescence was not ob-

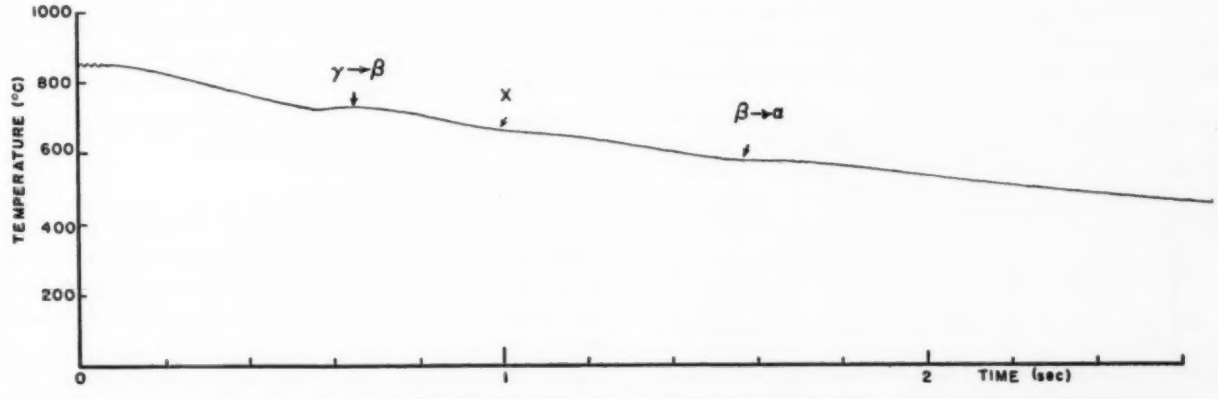


FIG. 2. Cooling curve obtained at a rate of 280°C/sec.

† Made by the William Miller Corporation, Pasadena, California.

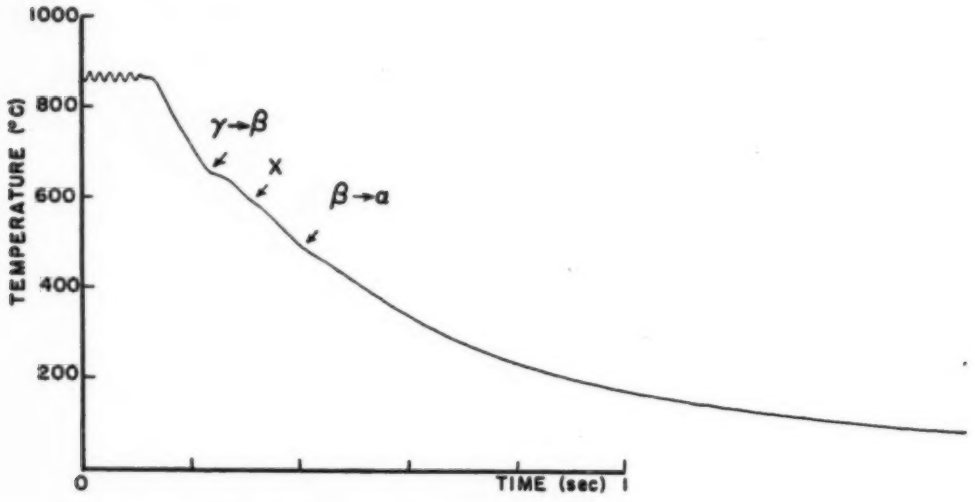


FIG. 3. Cooling curve obtained at a rate of 1800°C/sec.

served, the scatter is still greater than ±4°C, which is the estimated precision of measurement. It is probable that part of the variation in transformation temperature from specimen to specimen is due to the statistical nature of the nucleation process. In any case, all the measurements seem to be included in a band which has a width of about 30°C. By considering the center line of such a band as an average curve, it may be concluded that the gamma-to-beta transformation and the beta-to-alpha transformation temperatures of uranium decrease with increasing rates of cooling, and that the decrease is about 150°C for the gamma-to-beta transformation and about 250°C for the beta-to-alpha transformation, at cooling rates of the order of 8000°C./sec

As indicated on the graph of Fig. 4, the lowest rate of cooling used in the present study was about 5°C/sec. This rate is still much higher than that used in previous studies, from which the data presented in Table I were collected. In reference 11, for example, the rate of cooling is stated to be 1°C/min, or 0.0167°C/sec. The transformation temperature reported in reference 11 (765°C for gamma-to-beta and 655°C for beta-to-alpha), if plotted on the diagram of Fig. 4, would be two log cycles to the left of the vertical line $dT/dt=1$°C/sec. It is probable that between the rate of 0.0167°C/sec and the first point on the diagram of Fig. 4, obtained at a rate of 5°C/sec, the transformation temperature of both gamma-to-beta and beta-to-alpha decrease continuously with increasing rates of cooling.

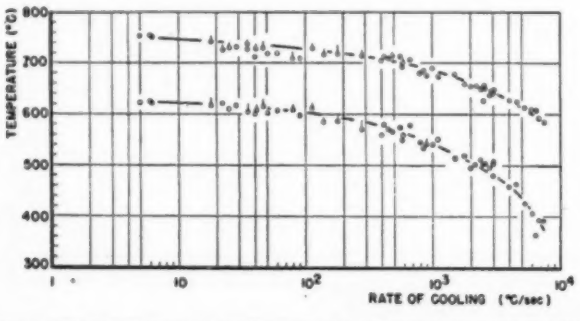


FIG. 4. Effect of rate of cooling on the allotropic transformation temperatures of uranium.

ANOMALOUS THERMAL ARREST WITHIN THE BETA-RANGE

In most of the recorded cooling curves (41 of 52 curves), a small but quite definite thermal arrest was noticed in the range of temperature between the two main arrests corresponding to the gamma-to-beta and the beta-to-alpha transformations (see point marked X in Figs. 1, 2, and 3). This extra thermal arrest was observed both in tests in which chromel-alumel thermocouples were used, as well as those in which platinum-platinum rhodium thermocouples were used. It was also present in experiments made by heating a specimen for the first time from room temperature into the gamma-field. A typical temperature-time curve obtained in these tests is presented in Fig. 5.‡

The extra thermal arrest was detected in cooling curves recorded at cooling velocities covering the entire range investigated. Within a certain scatter, which is not greater than that obtained for the two main transformation temperatures, it is apparent from the graph of Fig. 6 that the decrease in temperature of the extra thermal arrest with increasing velocity is very similar to that observed for the gamma-to-beta and the beta-to-alpha transformation temperatures (see Fig. 4).

RESULTS OF MEASUREMENTS ON HIGHER PURITY URANIUM

The results of measurements made with the higher purity uranium are summarized in Table II. For these tests, the rates of cooling were chosen so as to cover the range from 10 to 6500°C/sec. In all the cooling curves, the small thermal arrest within the beta-field was as

‡ It is interesting to note that upon heating at a rate of approximately 75°C/sec (see Fig. 5), a marked superheating effect occurs at the alpha-to-beta transformation temperature. The transformation temperatures are about 666 and 771°C, respectively. The agreement with the results obtained at a rate of 1°C/min is satisfactory (see Table I, reference 11).

clearly defined as on the curves obtained with the less pure metal. Furthermore, no significant differences were found between the transformation temperatures of the two kinds of uranium. It can be shown, indeed, that if the data given in Table II were plotted on the graphs of Figs. 4 and 6, the measured transformation temperatures of the higher purity uranium would not deviate from the previous data by more than about ±15°C. From the large number of results shown in Figs. 4 and 6, it seems that the scatter between individual measurements on the same material is also approximately ±15°C.

DISCUSSION

The results of the present investigation show that both the gamma-to-beta and the beta-to-alpha transformation temperatures are lowered when the cooling rate is increased. This result answers a question recently

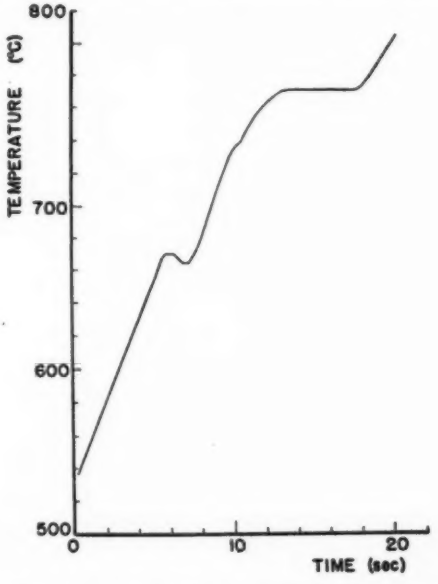


FIG. 5. Heating curve obtained at a rate of 75°C/sec.

raised by Pfeil[12] concerning the possibility of a direct gamma-to-alpha transformation under conditions of very fast cooling. The present findings indicate that such a direct transformation is very improbable, since (at least for rates of cooling up to about 8000°C/sec) the range of temperature in which the beta-phase exists seems to increase rather than decrease with the cooling rate.

The effect of the cooling rate on the transformation temperatures of uranium is very similar to that observed in iron, for which the gamma (face-centered cubic) to alpha (body-centered cubic) transformation temperature may be depressed by approximately 175°C when the rate of cooling is about 10 000°C/sec. In the case of iron, also, recalescence has been observed in many cooling experiments.[15] In contrast with the behavior of

[15] P. Duwez, Trans. Am. Inst. Mining Engrs. Met. **191**, 765–771 (1951).

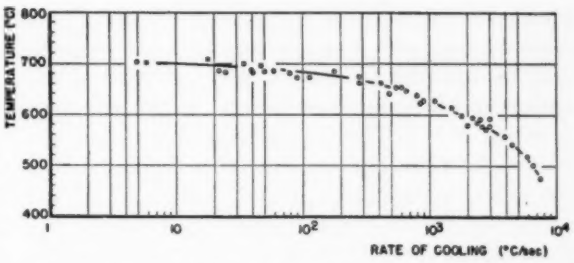


FIG. 6. Effect of rate of cooling on the temperature of thermal arrest observed within the beta-field.

both uranium and iron, the allotropic transformation in the four metals having a body-centered cubic high temperature phase and a hexagonal close-packed low temperature phase (titanium, zirconium, hafnium, and thallium) has been found to be very little affected, if at all, by the rate of cooling, and no recalescence seems to take place. For titanium and zirconium,[15] a slight decrease of the order of 15 to 25°C has been observed, and for hafnium[16] and thallium[15] no measurable decrease in temperature was detected. Although all these allotropic transformations are probably diffusionless, the degree to which they are affected by the rate of cooling may indicate essential differences in the mechanism of nucleation and shear through which the transformation takes place.

The rapidity at which the high temperature gamma-phase of uranium transforms into beta and into alpha, as evidenced by the present study, suggests that the atomic displacements required for the transformation must be rather small. So far, no atomic mechanism comparable to that proposed by Burgers for the transformation in zirconium[17] has been advanced for uranium. Now that the crystal structure of beta-uranium is known, such a mechanism will probably be uncovered with the aid of a study of the crystallographic relationships among the three phases.

The existence of a thermal arrest, however small in comparison with those corresponding to the two well-established allotropic transformations in uranium, suggests at first the presence of a new allotropic transformation. A thermal arrest, however, is not enough

TABLE II. Results of measurements on high purity uranium.

Rate of cooling (°C/sec)	Transformation temperature (°C) gamma-to-beta	X	beta-to-alpha
10	735–756[a]	...	...
27	737–749	708	...
40	730–744	690	595–602
54	716–728	682	598–612
60	730–745	694	610–620
85	730–740	682	600–615
117	702–716	661	590–598
325	715–720	661	582
2500	640	586	496
6500	585	497	405

[a] Two temperatures are given when recalescence was observed.

[16] P. Duwez, J. Appl. Phys. **22**, 1174–1175 (1951).
[17] W. G. Burgers, Physica **1**, 561 (1934).

evidence for establishing the existence of an allotropic transformation; and, if such a transformation existed, it would be difficult to explain why previous work on uranium and uranium-base binary phase diagrams did not reveal it.[9–12] The presence of impurities in the metal is of course always a possible argument for explaining any anomalous behavior. Although two grades of uranium were used for the present experiments, it may well be that the concentration of the critical impurity in the two grades was exactly the same and consequently produced the same effect. A third possible explanation for the thermal arrest recorded in the beta-temperature range might be based on the existence of a transition similar to that recently found by Fine *et al.* in chromium.[18] Such a transition occurring at 37°C in chromium is characterized by a rather abrupt change in Young's modulus, internal friction, coefficient of espansion, electrical resistivity, and thermoelectric power, without any change in crystal structure. Further experimental work on uranium is obviously necessary before any of the preceding hypotheses may be considered as a satisfactory explanation of the observed thermal arrest.

[18] Fine, Greiner, and Ellis, Trans. Am. Inst. Mining Met. Engrs. **191**, 56–58 (1951).

CONCLUSIONS

The transformation temperatures of both gamma-to-beta and beta-to-alpha uranium decrease continuously with increasing rates of cooling, over a range from 5 to 8000°C/sec.

A small but definite thermal arrest occurs between the two main arrests, which correspond to the two known transitions. This additional arrest was observed in most of the cooling curves and also in a heating curve.

Additional experimental work is required to establish whether the observed arrest is due to the presence of a hitherto unreported allotropic transformation, to chemical impurities in the uranium, to the type of transition recently found in chromium,[18] or to some other cause.

JOURNAL OF APPLIED PHYSICS VOLUME 24, NUMBER 2 FEBRUARY, 1953

Plastic Electrets

H. H. Wieder and Sol Kaufman
National Bureau of Standards, Washington, District of Columbia
(Received September 2, 1952)

A brief discussion of proposed explanations for the electret effect is presented. Plexiglas, Lucite, and Nylon electrets were prepared under fields ranging from 19 to 36 kv/cm and observed for as long as 2000 hours. The surface charge densities were measured by electrostatic induction using a commercial electronic electrometer and a shunt capacitance to reduce the readings to a maximum of 20 volts. The initial state was a heterocharge or a homocharge depending on the strength of the forming field. The steady state was a homocharge in every case. Charge densities as high as 5.5×10^{-9} coulomb/cm^2 (16.5 esu) were observed. A Plexiglas disk was charged at room temperature under a strong field and its subsequent decaying homocharge recorded. The experimental results support the ideas of Mikola and Gross, i.e., the existence of two decaying polarizations of opposite sense due to ionic migration, one occurring within the dielectric, the other, across the electrode-dielectric interface.

I. INTRODUCTION

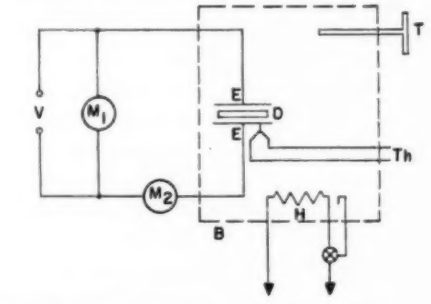
THE purpose of the studies described in this paper is to examine the properties of plastic electrets in the light of published information on carnauba wax electrets[1] and to clarify the basic mechanisms involved in fabrication and subsequent behavior of electrets. Relatively few investigations have been made on plastics,[2] although it is known that some exhibit the electret effect.

An electret is a dielectric which can maintain a sensibly permanent external electric field. It is usually prepared by heating and subsequent cooling of a suitable material while subjected to a high potential. Although many substances can be electrified at room temperature by means of friction or through the application of strong electric fields, the resulting surface charges tend to decay rapidly. If these substances are immersed in water or exposed to humid or highly ionized air, their charges will be irreversibly neutralized. Electrets, on the other hand, when properly stored will not decay and if neutralized as in the above manner can recover to full strength.

Many explanations for the electret effect have been proposed. Adams[3] speaks of a pyroelectric effect with a long relaxation time. Gemant[4] attributes the external field to two opposing polarizations. One arises from real ionic charges which under the forming field have concentrated at the surfaces of the electret; these

[1] See extensive bibliography in F. Gutmann, Revs. Modern Phys. **20**, 457 (1948).
[2] F. Binder, Z. Naturforsch. **6a**, 714 (1951).
[3] E. P. Adams, J. Franklin Inst. **204**, 469 (1927).
[4] A. Gemant, Phil. Mag. **20**, 929 (1935); Elect. Eng. **68**, 644 (1949).

charges are opposite in sign to the charges on the adjacent forming electrodes and are thus called heterocharges. The other is dipole orientation of the dielectric by the field of the ionic heterocharges. This new polarization is representable by fictitious surface charges of the same sign as the charges on the adjacent electrodes and are called homocharges. The dipole orientation remains frozen, while the real ionic heterocharge decays. Mikola[5] and later Gross[6] also postulate two distinct charge densities; one is due to ionic charges which occur through breakdown between the electrode and dielectric; the other is due to polarization within the dielectric. Using Mikola's terminology, these processes may be referred to as outer polarization and inner polarization, respectively. The former gives rise to homocharges, the latter to heterocharges. Both polarizations may be large. The net surface charge density is determined by the difference between the two polarization magnitudes and could be considerably smaller



V- HIGH VOLTAGE SUPPLY-VARIABLE 0-10 KV
M_1- VOLTMETER
M_2- ELECTRONIC MICROAMMETER
E- BRASS ELECTRODES
D- DIELECTRIC
B- TEMPERATURE REGULATED CHAMBER
T- INDUSTRIAL THERMOMETER
Th- THERMOCOUPLE
H- HEATER

FIG. 1. Experimental set-up for electret preparation.

in value. A potential barrier and high internal resistivity will cause both polarizations to decay slowly, leaving a net homocharge. The actual mechanism of inner polarization is thought to be an ionic space charge set up within the dielectric at the forming temperature and frozen in when room temperature is restored. Investigations by Thiesen *et al.*[7] indicate that in the case of a carnauba wax electret there exists an internal space charge. They ascribe this to charges being transferred from electrodes to the dielectric and ion migrations through the wax. Measurements on a surface charged dielectric formed at high field strength and at room temperature (homocharge) indicated that the volume distribution of charge is negligible except in a layer immediately adjacent to the surfaces. By superimposing space charge measurements from two samples, one prepared in the above described manner (homocharge) and the other having a predominant heterocharge, Thiesen found a correlation between the resultant curve and the space charge distribution curve of a sample which had undergone a sign reversal from heterocharge to homocharge in the typical electret manner. These measurements by Thiesen, as well as our experimental results, tend to corroborate the ideas of Mikola and Gross. The mechanism of outer polarization has been thoroughly investigated by Gross.[6,8] Swann[9] uses a macroscopic approach which assumes a semipermanent polarization and a real surface and volume charge density. He arrives formally at a result which describes many aspects of electret behavior. This development is compatible with the theories of Mikola and Gross.

[5] S. Mikola, Z. Physik **32**, 476 (1925).
[6] B. Gross, Phys. Rev. **66**, 26 (1944); J. Chem. Phys. **17**, 866 (1949); B. Gross and L. F. Denard, Phys. Rev. **67**, 253 (1945).
[7] Thiesen, Winkel, and Hermann, Physik. Z. **37**, 511 (1936).

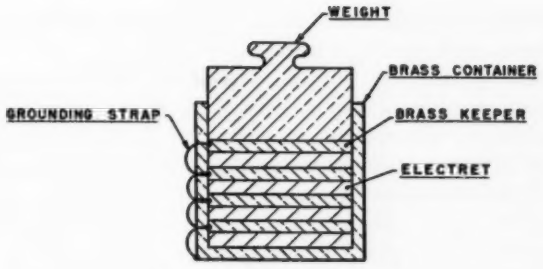
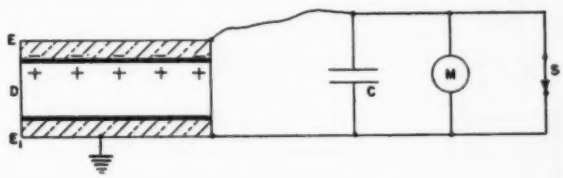


FIG. 2. Electret storage unit.

II. EXPERIMENTAL PROCEDURE AND RESULTS

The subject of investigation was the behavior of electrets and surface charged dielectrics of Plexiglas, Lucite (both polymethylmethacrylates), and Nylon (polyhexamethylene adipamide). The experimental method used in preparing the electrets is shown in Fig. 1. The temperature of the box was gradually brought up to approximately 150°C in the case of Plexiglas and Lucite and to 200°C in the case of Nylon. The total current *versus* time was recorded, and when the former reached a steady-state value the box was cooled to



M- ELECTROMETER- KEITHLEY MODEL 200
C- 0.035 μfd. POLYSTYRENE CAPACITOR
S- SHORTING SWITCH
E- MOVABLE BRASS ELECTRODE
E_1- FIXED ELECTRODE
D- ELECTRET

FIG. 3. Schematic of circuit for charge density measurement.

[8] B. Gross, Brit. J. Appl. Phys. **1**, 259 (1950).
[9] W. F. G. Swann, J. Franklin Inst. **250**, 219 (1950).

ambient temperature (24°C). Throughout this time the applied field was kept constant.

After formation, the electret was stored in a container shown in Fig. 2. This effectively shorts the active surfaces of the electret and thus confines the external field to the minute volume between the conducting and dielectric surfaces; this prevents decay arising from outside agents. No attempt was made to control the humidity, temperature, and pressure during storage as was done by previous experimenters using carnauba electrets.

The method of electrostatic induction was used for measurements of the surface charge density. The apparatus is shown in Fig. 3, and the following experimental method was used. Electrode E is first grounded by means of switch S. Upon opening S and lifting E, the potential across C is read on the electrometer. With E now in the lifted position, S is again closed returning E to ground potential. Reopening S and lowering E to its original position result in an electrometer reading opposite and nearly equal to the first reading. The magnitudes are averaged and combined with the sign of the second reading to obtain the quantity $\bar{V}$. If A is the area of electrode E, Q is the total charge induced on E when E is brought from the lifted to the resting position and σ is the surface charge density; then $\sigma=Q/A=C\bar{V}/A$. This process is repeated for each surface of the electret. Loading the electrometer with a 0.035 μf capacitance reduced the measured potentials to the order of 10 volts. Measurements made at higher potentials would have been less precise due to increased leakage and corona effects.

(A) A Plexiglas disk, 10.2 cm in diameter and 0.14 cm thick, was made into an electret under a forming potential of 5 kv ($E=36$ kv/cm). Figure 4 indicates the behavior during formation.

The current curve may be resolved into a steady-state conduction current and a transient polarization current. The latter refers solely to inner polarization, as the outer polarization involves only an infinitesimal displacement of charge across the electrode-dielectric interface and, hence, a negligible external current. Integration of the polarization current yields a total charge of approximately 4×10^{-2} coulomb or an effective surface heterocharge density of 5×10^{-4} coulomb/cm². However, Fig. 5 shows that when the electret's net surface density is measured after formation, a homocharge in the order of 3×10^{-9} coulomb/cm² is obtained; this is several orders of magnitude smaller and of opposite polarity. A simultaneously occurring outer polarization large enough to mask the inner polarization could account for the measured net charge density. The variation in charge density with time can be explained by the interplay of two mechanisms: one being a differential in the decay rate of the hetero- and homocharge which tends to increase the net homocharge, the other being the sporadic and localized breakdowns of the surrounding air which tend to decrease the net homocharge. The precise point of breakdown is probably dependent on a number of as yet uncontrolled factors such as ambient temperature, humidity, barometric pressure, and degree of the separation between electret and keeper surfaces.

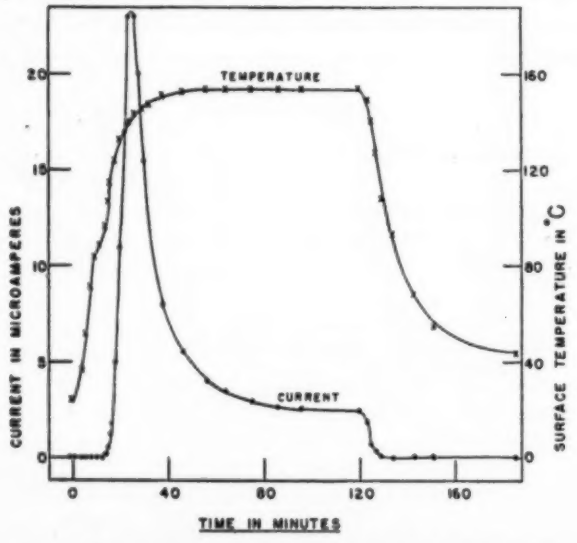


FIG. 4. Current during formation of Plexiglas electret (A).

(B) A Lucite disk 10.2 cm in diameter and 0.32 cm thick was made into an electret under a forming potential of 6 kv ($E=19$ kv/cm) and a temperature of 155°C. The time plot of forming current was similar to that shown in Fig. 3 except for a broadening of the transient phase; this was probably due to the increased time required for a thicker disk to reach temperature equilibrium. The surface charge density variation with time is shown in Fig. 6.

This electret started with a heterocharge which decayed, changed sign, and increased to a relatively static homocharge. Note that the use of a weaker forming field gave the electret an initial heterocharge which can be explained as follows. The reduced value of the forming field prevented any outer polarization from occurring until the inner polarization had attained a sufficiently high value to increase the total field strength in the electrode-dielectric interface. This excess of inner polarization over outer polarization was then maintained until formation was complete. The differential in the decay rates of the two polarizations caused the charge reversal and subsequent build-up of a net homocharge. The opposing mechanism of air breakdown which limits the homocharge made its appearance at a charge density level of about 3×10^{-9} coulomb/cm². The discontinuity at 300 hours in Fig. 6 was due to deliberate immersion of the electret in water. Additional electrets of Plexiglas and Lucite have been prepared, and results paralleling those described above have been obtained.

(C) A Lucite electret 10.2 cm by 0.32 cm was prepared under an even smaller forming field ($E=6$ kv/cm). As expected a heterocharge initially appeared (Fig. 7). A

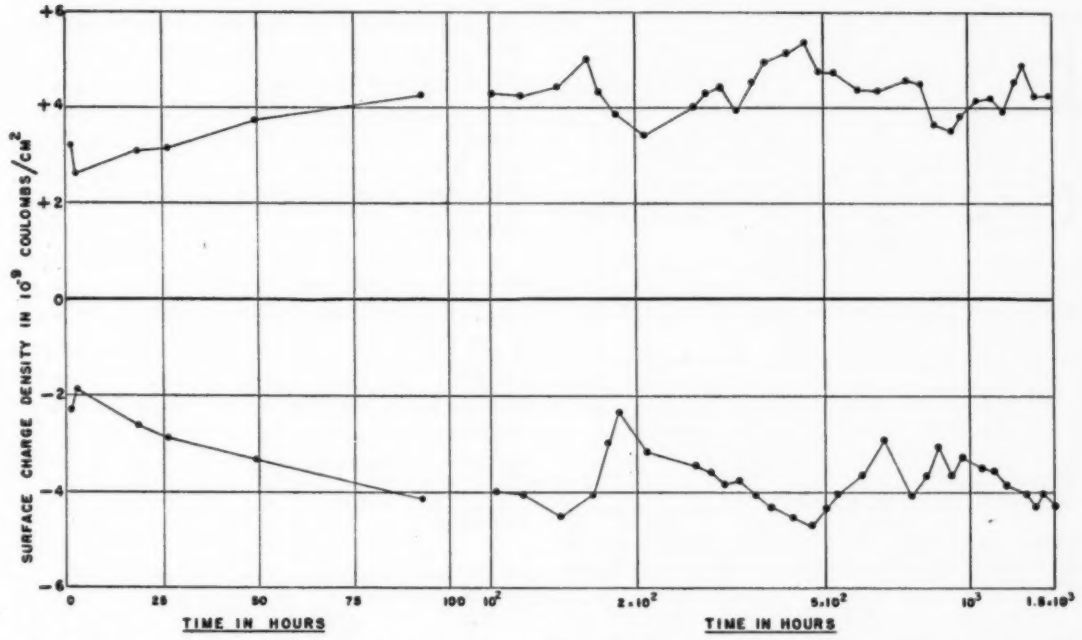


FIG. 5. Surface charge density *vs* time for both surfaces of Plexiglas Electret (A). Homocharge with no sign reversal.

tendency toward charge reversal is apparent. The smaller $d\sigma/dt$ is attributed to a smaller total inner and outer polarization with a resulting decreased differential in homocharge and heterocharge decay. After 125 hours the electret was subjected to a field of 30 kv/cm for about one minute. This produced some additional outer polarization which shifted the net charge density to a homocharge. The low rate build-up of homocharge continued. A similar experiment by Thiesen *et al.*[7] was performed on a carnauba wax electret.

(D) In order to test the behavior of a surface-charged dielectric as compared to an electret, a Plexiglas disk 10.2 cm diameter and 0.14 cm thick was subjected to a field of 53 kv/cm for approximately five minutes at room temperature. Figure 8 shows the decay with time of the resulting homocharge on the Plexiglas surfaces. Again agreement was noted with similar results obtained by Thiesen.[7]

(E) Measurements of space charge distribution on plastic electrets such as those performed by Jaeger[10] and Thiesen[7] were not practical in this case. However, a qualitative test was performed as follows: A 10.2-cm diameter and 0.15-cm thick Plexiglas electret which had attained a steady homocharge was reduced to a thickness of 0.12 cm by machining the positively charged surface in a lathe. Care was exercised to insure that a minimum amount of heat be generated during the operation. A decrease in charge density occurred from approximately 4.7 millimicrocoulomb/cm² to 1.4 millimicrocoulomb/cm² just after machining. A gradual recovery of the charge density to a steady-state value of approximately 2.7 millimicrocoulomb/cm² ensued. The existence of a homocharge interior to the original surface indicates that the charges associated with the outer polarization have penetrated the dielectric. This penetration may have occurred during the high temperature, low resistivity, phase of electret formation.

(F) A Nylon disk 5.1 cm in diameter and 0.16 cm thick was made into an electret under a forming field of 25 kv/cm at a temperature of 200°C using the experimental setup described before. The current during formation (Fig. 9) indicated no transient such as that shown for Plexiglas (Fig. 4) or Lucite. The total current increased rapidly with temperature and thus attained a high conduction level which masked any accompanying polarization current. An initial net heterocharge was noted (Fig. 10) which reversed to a homocharge,

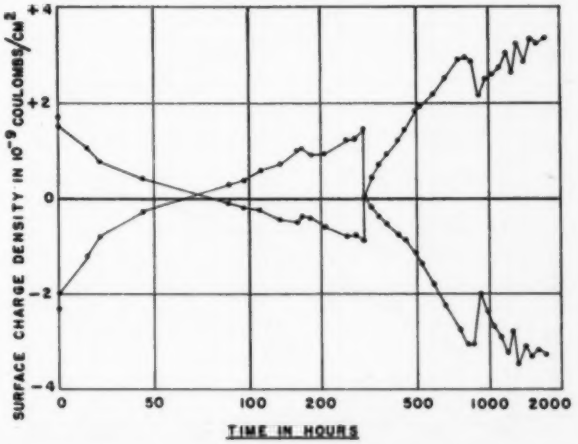


FIG. 6. Surface charge density *vs* time for both surfaces of Lucite electret (B). Initial heterocharge, reversal to homocharge at 70 hours, recovery of homocharge after water immersion at 300 hours.

[10] P. Jaeger, Ann. Physik **21**, 481 (1934).

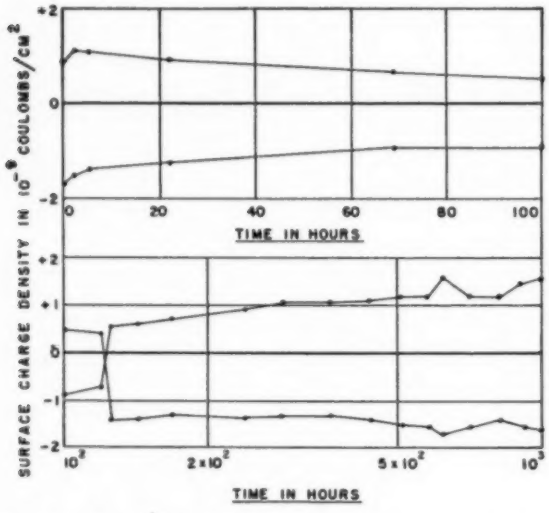


FIG. 7. Surface charge density *vs* time for both surfaces of Lucite electret (C). Initial heterocharge, forced reversal to homocharge at 125 hours.

the reversal in sign occurring in a somewhat shorter time than in the case of Lucite. This indicates a larger differential in the decay rates of the inner and outer polarizations. The magnitude of the surface charge density is again limited to a value approximately equal to that of the Plexiglas samples due to the previously described breakdown of the air.

III. SUMMARY AND CONCLUSIONS

1. Electrets have been successfully prepared from Plexiglas, Lucite, and Nylon. The current curve during formation of the Plexiglas and Lucite electrets is similar to that observed in the case of carnauba wax. In the case of Nylon, however, the formation current differs in shape because of a higher conduction current; this in turn masks any polarization current which might be present. Although no over-all criteria for selecting suitable electret materials have been established, a high volume resistivity (greater than 10^{14} ohm cm) seems to be essential.

2. Field strengths of the order of 35 kv/cm employed during formation result in electrets having an initial net homocharge; if the field strength is in the range of 20 to 25 kv/cm the resulting electret will have an initial heterocharge which subsequently becomes a homocharge. If the forming field is below 6 kv/cm, a slowly decaying heterocharge is observable. The highest detectable charge density is limited to an order of magnitude of 10^{-9} coulomb/cm^2 by breakdown of the surrounding air.

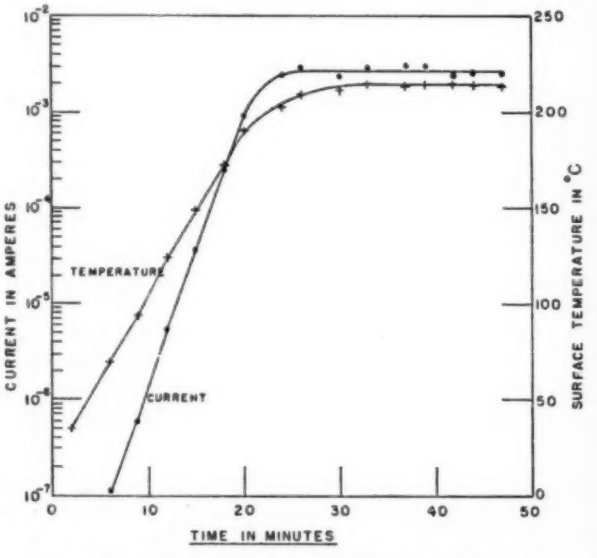


FIG. 9. Current during formation of Nylon electret (F).

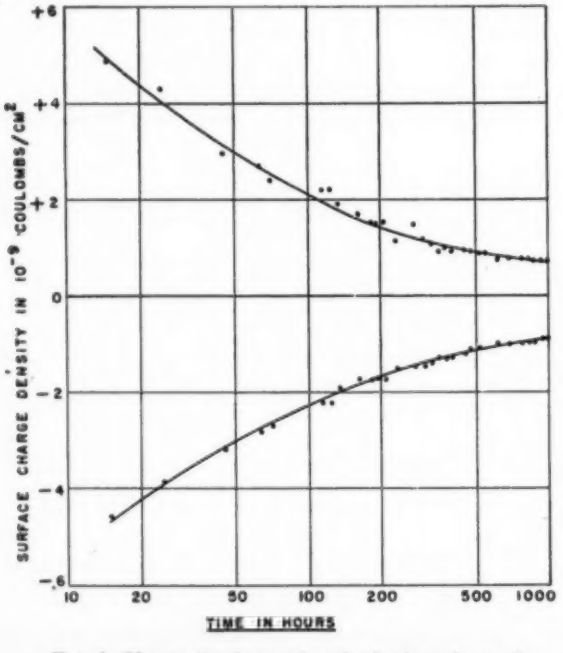


FIG. 8. Charge density *vs* time for both surfaces of surface charged Plexiglas disk (D). Homocharge.

3. Since the process of growth of a simple polarized charge distribution is unlikely, it is reasonable to postulate, in agreement with Mikola[5] and Gross,[6] the existence of two separate, decaying charge distributions of opposite polarities. A differential in their decay rates could result in the apparent growth of the total charge density in the direction of a greater homocharge. This growth is eventually limited and an approximate steady state reached because of air breakdown. Both polarized distributions involve the mechanism of ionic charge displacement. The alternative possibility that dipole orientation is the basis of inner polarization is unreasonable since the maximum surface charge density obtainable in this way (assuming complete alignment with reasonable magnitudes for dipole moment and dipole density) is 5×10^{-6} coulomb/cm^2. In contrast the actual charge density due to inner polarization for electret *A* was computed by integrating the displace-

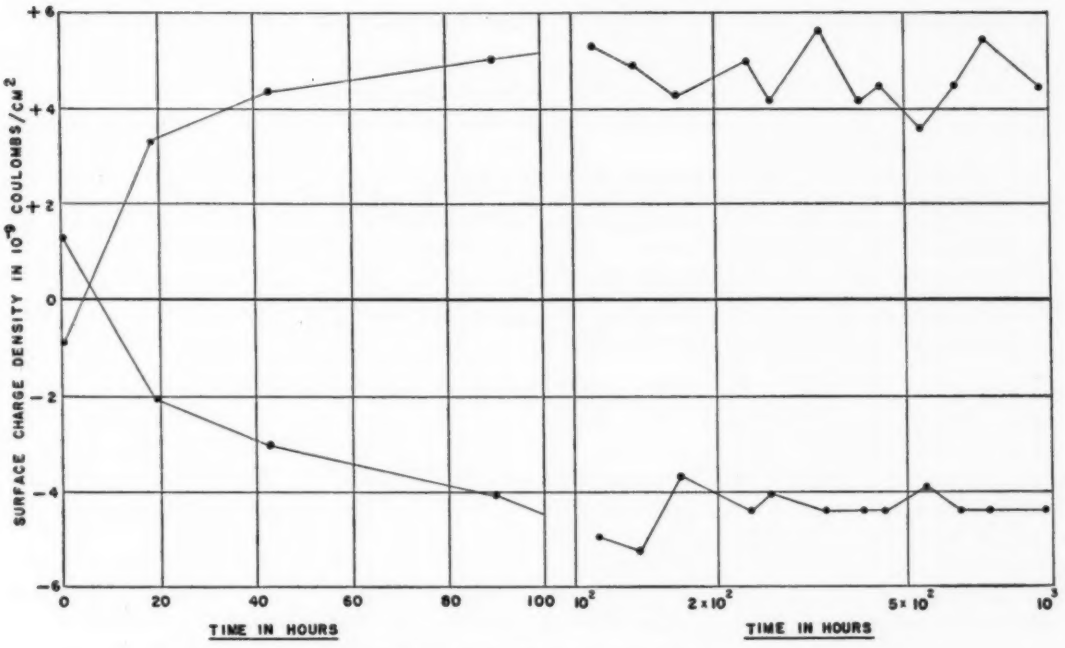


FIG. 10. Surface charge density *vs* time for both surfaces of Nylon electret (*F*). Initial heterocharge, reversal to homocharge at 8 hours.

ment current-time curve of Fig. 4 and found to be 5×10^{-4} coulomb/cm^2 (assuming conductivity to be independent of time).

4. Substances having a slowly decaying heterocharge (Fig. 7) could be electrets which, because of a small differential in decay rates, require a long time for reversal of sign and build-up of homocharge. A second possibility is that they have a pure heterocharge density due to inner polarization alone and thus can only decay to zero.

5. Surface charged dielectrics prepared under a field of about 55 kv/cm at room temperature exhibited a slowly decaying homocharge (Fig. 8). The absence of elevated temperatures prevented inner polarization and thus led to a pure homocharge which decayed directly toward zero. A similar situation would probably arise in an attempt at electret formation using a material which is incapable of appreciable inner polarization even at higher temperatures. Note that nonelectrets, i.e., dielectrics having a pure heterocharge or a pure homocharge may have their rates of decay reduced sufficiently after a considerable period of decay so as to appear sensibly permanent.

6. Nylon electrets in contrast to Plexiglas and Lucite have a greater activity as evidenced by the former's more rapid sign reversal and attainment of steady-state homocharge. This can be caused by the presence of a larger inner and outer polarization or a lower volume resistivity in the Nylon. As yet only the latter is known to be true.

7. The ability of electrets to recover from external neutralization is demonstrated by the Lucite electret of Fig. 6.

8. The attempts to ascertain the actual charge distributions within electrets by shaving off of successive layers or by other contemplated techniques are of a preliminary nature and not, as yet, interpretable.

Several areas of activity for future work are suggested here. An improved technique in space charge measurements within solid dielectrics would be useful toward constructing an adequate model for the electret. Along this vein, tentative experimentation and study of film electrets is an unexplored area. An examination of a wider class of plastic materials should result in a greater variety of available electrets and help to establish empirical criteria which determine whether or not a given material can be made into an electret. This will be the subject of a future report.

The authors are indebted to Dr. John Moorhead (National Bureau of Standards Staff) for aid and advice and to Mr. William Giokas for the preparation of the charts and drawings.

A Thermoelectric Study of the Electrical Forming of Germanium Rectifiers

MAKOTO KIKUCHI AND TOMIO ONISHI
Electrotechnical Laboratory, Tokyo, Japan
(Received November 25, 1952)

The electrical forming phenomena on the etched surfaces of *n*-type Ge crystals by ac voltages were studied thermoelectrically. Remarkable improvements in the i–v characteristics were obtained by applying appropriate forming voltages. The thermoelectric current observed on the etched surface showed *n*-type direction but it was converted to *p*-type by the forming. It has also been found out that the thermo-emf depends sensitively on the pressure of the whisker contact after the forming (and on the polished surface), while the dependence is small before the forming on the etched surface.

Experimental results seem to show that some substance of relatively high resistivity and of *p*-type thermo-emf is produced by the forming between the whisker metal and the Ge surface.

I. INTRODUCTION

IT was found out by Granville and Hogarth[1] that the thermo-emf at the point contact on etched Ge surfaces, which is *n*-type as expected from its bulk properties, could be converted to *p*-type by polishing the surface. We have attempted to apply this experimental method to the forming phenomena of Ge rectifiers.

Several experimental results have been reported on the electrical forming but some of them disagree with each other. For instance, Bardeen and Pfann[2] have stated that the results of the forming depend on the direction of the passage of the forming current, i.e., the effect is due to the migration of impurity centers near the contact by the *electric field* produced by the forming current. On the other hand, Stelmak[3] and Longini[4] have shown that the results *do not* depend on the direction of the forming voltages and that the effect is essentially due to the *heating* of the contact. We can not assert at present which of the above two processes acts in reality.

We have used ac forming voltages to eliminate the effect of the electric fields such as indicated by Bardeen and Pfann and have attempted to observe the changes of the thermoelectric current and rectifying characteristics by the forming.

II. EXPERIMENTAL METHOD

The circuits used in our experiments are shown in Fig. 1. Circuit 1 is for measuring the dc characteristics, circuit 2 is a galvanometer to measure the thermoelectric current, and the electrical forming is performed through circuit 3. Figure 2 shows the construction of the heater in detail. Manganin wire of 0.1 mm diameter is wound on a thin insulating tube through which the tungsten whisker is passed. These are fixed on a support which is movable on a vertical line. The end of the tungsten wire is pointed by electrolytic etching.

One side of the Ge specimen was plated with Ni and was soldered on a copper plate, which was mounted on a micromanipulator to be moved in two perpendicular directions in a horizontal plane. The resistivity of the Ge was 0.027 ohm-cm at room temperature.

The temperature of the heater *H* was controlled by adjusting the output voltage of a Slidac. The temperature of the tip of the tungsten whisker was previously calibrated as a function of the output voltage of the Slidac using a thermocouple.

In the following experiments, the temperature was kept constant at 16°C which was 4°C above the room temperature. When the measurements were going on, the contact point was covered by a paper cap to prevent the influence of the movements of the air (for example caused by breathing) since they had considerable effects to cool the contact.

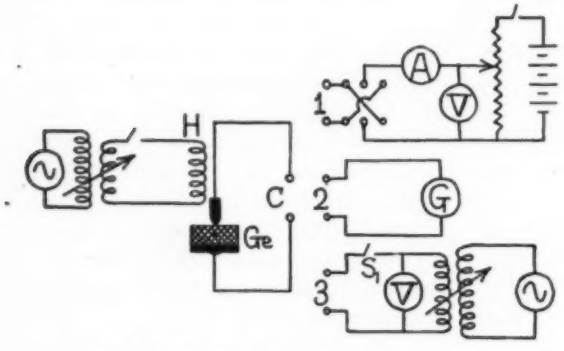


FIG. 1. The circuits used in this experiment.

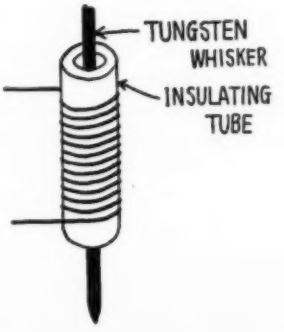


FIG. 2. The construction of the heater.

[1] J. W. Granville and C. A. Hogarth, Proc. Phys. Soc. (London) 64, 488 (1951).
[2] J. Bardeen and W. G. Pfann, Phys. Rev. 77, 401 (1950).
[3] J. P. Stelmak, Phys. Rev. 83, 165 (1951).
[4] R. L. Longini, Phys. Rev. 84, 1254 (1951).

III. EXPERIMENTAL RESULTS

A. Inversion of the Thermoelectric Current by the Forming

The thermoelectric current is kept *n*-type after the forming for low forming voltages, but it becomes more probable that it will be converted to *p*-type as the forming voltages are increased, and this probability approaches unity when the forming voltage exceeds 8 volts.

It has been found out that the direction of the thermoelectric current after the forming depends on the pressure of the whisker contact as shown in Fig. 3. *P*-type, thermoelectric current could be observed by decreasing the pressure on the whisker even if the thermoelectric current after the forming was *n*-type. Similar effects as shown in Fig. 3 were observed on the polished surface.

B. Investigation of the Substance Produced on the Surface

After forming a point on the etched surface by considerably large forming voltage, we lifted the whisker

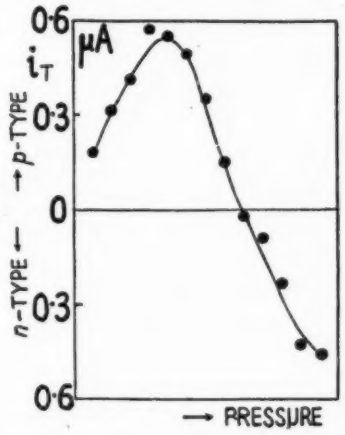


FIG. 3. Variation of the thermoelectric current with the pressure of the whisker contact after the forming.

and put it on another point on the same surface. Then *p*-type thermoelectric current was observed. This *p*-type current, however, was converted to *n*-type again by grinding the tip of the whisker by an oilstone.

Next, after forming a point, we ground the tip of the whisker and put it again on the place where the forming had been performed. Then *p*-type thermoelectric current was observed.

These phenomena show that some substance of *p*-type thermo-emf is produced at the contact by the forming, which can be carried sticking on the tip of the whisker.

To test whether this substance might be an oxide of Ge, the following experiments were performed. Ge smoke, evaporated from an arc discharge between two Ge electrodes, was condensed on an etched copper plate and a tungsten whisker was put on it as shown in Fig. 4. *P*-type thermoelectric current was found on this substance. (Any indication of *p*-type thermoelectric current was not found on the surface of the etched copper surface.)

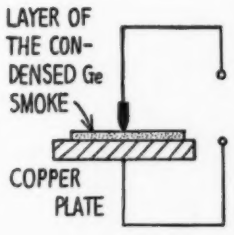


FIG. 4.

The i-v characteristics of this specimen are shown in Fig. 5. As shown in the figure, it is almost ohmic when the pressure on the whisker is comparatively large, but shows slight *p*-type rectification when the pressure is small.

The resistivity of this substance between the whisker and the copper plate was roughly estimated to be between 1 and 50 ohm-cm.

Similar experiments were performed by condensing the tungsten smoke, but we could not find out *p*-type thermoelectric current at all in this case. These results seem to compel us to believe that *p*-type substance is an oxide of Ge produced by the heating of the contact, and the tungsten whisker does not play an important role.

C. Equivalent Circuit and the Determination of the Thermo-emf from the Thermoelectric Current

The equivalent circuit as shown in Fig. 6 can be used to treat our experiments quantitatively. V_T is the thermo-emf produced at the contact, R_C the resistance of the rectifier as a whole, i_T the thermoelectric current, R_K is a variable resistance, and G the galvanometer. Resistances of negligible quantities such as the inner resistance of the galvanometer are not shown in this figure. Values of V_T and R_C which are enclosed by the broken line in the figure cannot be measured *directly* from this measurement, but we can determine them separately from the value of i_T for two different values of R_K using simple equations.

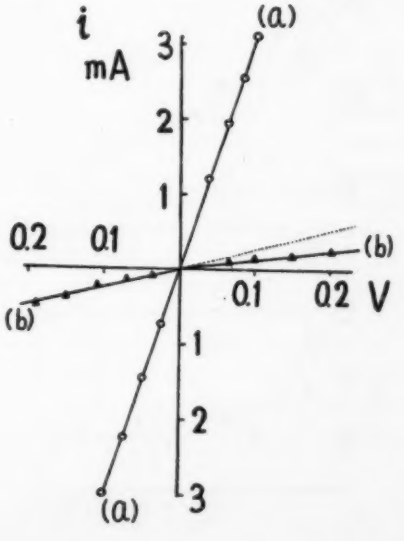


FIG. 5. i–v characteristics of the specimen shown in Fig. 4.

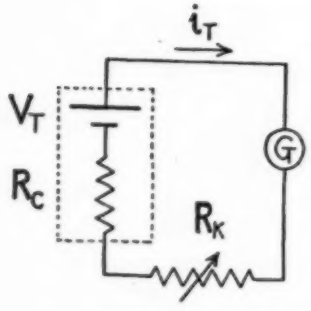


FIG. 6. Equivalent circuit.

As was expected, the values of R_C thus obtained were in good agreement with the low level resistance R_o of the rectifier measured at $f=1$ kc.

D. Variations of V_T and R_C

V_T and R_C of the point contact were measured by this method as a function of the pressure on the whisker. Typical results are shown in Fig. 7 and Fig. 8. Figure 7 shows the variations of V_T and R_C on the etched surface *before the forming*, and Fig. 8 shows them *after the forming*. Similar dependency of V_T and R_C on the pressure on the whisker was observed on polished surfaces but the values of V_T seemed to show that the thickness of the layer of p-type substance was much larger in this case.

The value of V_T on the etched surface decreases when pressure is far reduced as shown in Fig. 7. This might be an evidence of the existence of an extremely thin layer of p-type even on the etched surface at room temperature.

Figure 9 shows the variation of V_T and R_C by successive forming operations. Resistance increases by the forming and the thermo-emf tends to change its sign to p-type. The monotonic changes as shown in Fig. 9 occurs for comparatively low forming voltages but fluctuations come into this change for larger forming voltages.

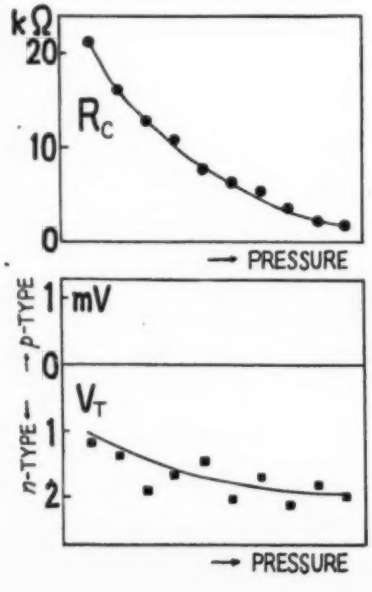


FIG. 7. V_T and R_C as a function of the pressure on the whisker on the etched surfaces before the forming.

F. Experiments on Si Crystals

The thermoelectric current observed on etched surfaces on Si crystals by the same method as mentioned above was p-type, and was not converted to n-type by polishing. This means that it is much more difficult to produce some substance layer on Si surface than on Ge by usual polishing techniques.

It was also very difficult to convert the direction of the thermo-emf to n-type by the forming. This change occurred only for very large forming voltages (>25 v) but with much less reproducibility than on Ge.

G. Improvement of the Rectifying Characteristic by the Forming

The rectifying characteristic of a point contact on the electrolytically etched surface of our n-type Ge was improved remarkably by the forming with appropriate values of forming voltages. A result for the forming voltage of 4 v is shown in Fig. 10. The backward current i_B at 0.4 v is decreased to 1/100 of the initial value, whereas the forward current i_F at 0.4 v is almost unchanged. Such an improvement could not be obtained with both too low or too high forming voltages.

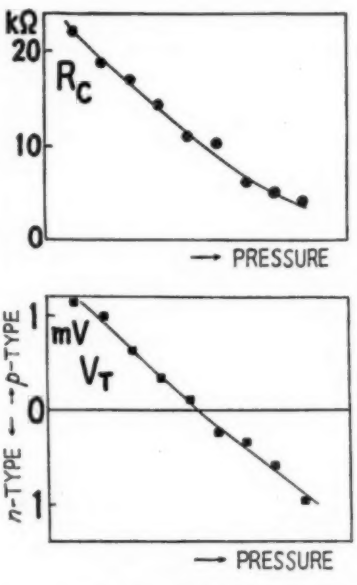


FIG. 8. V_T and R_C as a function of the pressure on the whisker on the etched surfaces after the forming.

It must be noted that such a large improvement of rectifying characteristic by the forming as obtained on the etched surface could not be found on the polished surface.

H. Other Experimental Results

It has been found out that there is a remarkable difference between a point contact on the etched surface and that on the polished surface in the dependence of the rectifying characteristic on the pressure of the whisker contact.

The rectification ratio i_F/i_B at 0.4 v remained almost constant on the polished surface, while on the etched surface it decreased when the pressure on the whisker was increased and vice versa.

The rectifying characteristic of the point contact was changed successively by the forming, but it has been found out that the low level resistance R_o and the backward resistance R_B measured at 0.4 v varies under the relation

$$R_B \propto R_o^n (n=1.2\sim1.3),$$

which holds in the range of $0.5\ k\Omega < R_o < 50k\Omega$, while the values of the forward current i_F at 0.4 v vary at most ± 20 percent and have no clear relation with these two quantities.

IV. DISCUSSION OF RESULTS

The results mentioned in Sec. III, A and D can be explained theoretically using comparatively simple model of the contact.

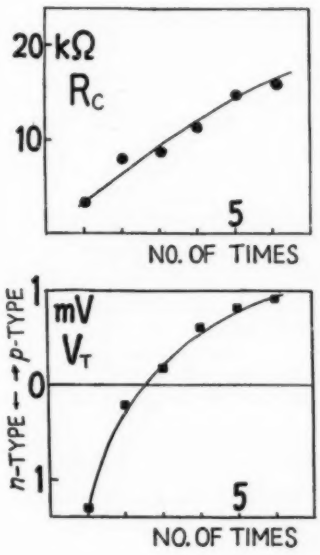


FIG. 9. Variation of V_T and R_C by the successive forming.

We assume that the magnitude of the thermo-emf, V_p, produced by a p-type substance, can be expressed for the first approximation as follows:

$$V_p = \alpha_p \Delta T_p, \tag{1}$$

where α_p is a constant of the p-type substance, and ΔT_p is the temperature difference across it. The same expression for an n-type substance can be used by changing the suffix p to n.

If we consider that there is a p-type layer of thickness D on the surface of n-type Ge crystal, we can calculate the magnitude of the total thermo-emf V_T as the difference of V_n and V_p, because the thermo-emf of the p-type layer and the n-type Ge acts in opposite direction.

Assuming that the temperature distribution can be replaced by straight lines, V_T is expressed as follows:

$$V_T = V_n - V_p = \alpha_n(T_a - T_o) - kD(\alpha_n + \alpha_p) \quad \text{for} \quad T_a \geqq T' \tag{2}$$

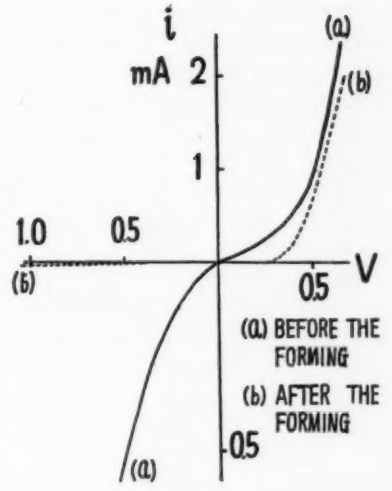


FIG. 10. A typical result of the improvement of the i–v characteristic by the forming.

and

$$V_T = -\alpha_p(T_a - T_o) \quad \text{for} \quad T_a < T', \tag{3}$$

where

$$T' = T_o + kD \tag{4}$$

and k is a phenomenologically introduced constant of the p-type layer. T_o and T_a are the temperature of the whisker point and of the base electrode, respectively. Equation (3) is for the case in which there is practically no temperature difference between the top and the base of the n-type Ge.

Considering that the effect of the pressure on the whisker acts through the thickness D and the contact area, the results shown in Sec. III, A and D can easily be explained by Eq. (2).

As for the results shown in Sec. III, G and H, existing "single contact" theories, e.g., Mott's theory and Schottky's theory, fail to explain them consistently, since the variation of parameters such as contact area A and the contact potential difference ϕ in the equations of these theories have influences of the same order of magnitude both in the forward and reverse directions.

One of the reasons why they are weak in explaining those results is that they do not take into account the mechanisms which are predominant in the reverse direction, whereas such mechanisms seem to have significant influences even on the low level characteristics. Theories which take them into account without changing the basic models of the barrier are not sufficient to explain the discrepancy of the inclination of the log i-v curve.[5]

These facts seem to show that the fundamental improvement of the "single contact" rectification theories must be done with changing *the model* of the contact. For this, the p-type layer on the n-type Ge which has been studied in our experiments will play an important role.

[5] E. D. Courant, Phys. Rev. **69**, 684 (1946).

We conclude that there must be a maximum in the rectification ratio for an appropriate value of the thickness of the surface layer. Too thin a layer is not enough to prevent the backward currents due to such effects as tunneling or image force. On the other hand, if there were too thick a layer on the surface, it might be expected that the current will flow in this layer along the surface, and thus the rectification ratio would be lowered.[6]

So-called "multicontact" theory is kept from failing through its arbitrary parameters. We cannot, however, conclude from this that the multicontact theory is the correct one because there still remains some essential problems about the "distribution function" of the contact potential difference.[7]

We propose that it is necessary to investigate the i-v characteristic using a new model which takes the p-type layer into account if we want to deal with the "single contact" theory, while some conditions on the "distribution function" of the contact potential difference must be found out when we are going to deal with the "multicontact" theory.

The authors wish to express their indebtedness to M. Hatoyama and the members of our semiconductor group, and to M. Shibuya, W. Sasaki, and Y. Tarui for valuable discussions.

[6] C. A. Hogarth and J. W. Granville, Proc. Phys. Soc. **64**, 992 (1951).

[7] H. J. Yearian, J. Appl. Phys. **21**, 214 (1950); Johnson, Smith and Yearian, J. Appl. Phys. **21**, 283 (1950).

JOURNAL OF APPLIED PHYSICS VOLUME 24, NUMBER 2 FEBRUARY, 1953

Hall Effect Modulators and "Gyrators" Employing Magnetic Field Independent Orientations in Germanium

W. P. Mason, W. H. Hewitt, and R. F. Wick
Bell Telephone Laboratories, Murray Hill, New Jersey

(Received August 7, 1952)

Three uses for the Hall effect in germanium crystals are described. These are (1) use of Hall effect probes in measuring magnetic flux, (2) use of Hall effect in crystals to produce a pure product modulator, and (3) use of Hall effect in germanium crystals to produce a nonreciprocal transmission. If the resistances are shunted around such gyrators, the transmission can be made zero in one direction and finite in the other. In all these applications, use is made of a crystal orientation for which the cross magneto-resistance effects are zero and the Hall effect constant does not vary with field by more than 2 percent out to a flux density of 20 000 gausses. This orientation was located by making a phenomenological study of the magneto-resistance and Hall effect corrections for a cubic crystal and evaluating constants experimentally. Correction terms to fourth and fifth powers of the magnetic field have been obtained.

I. INTRODUCTION

When current flows through a crystal of germanium in one direction and a magnetic field is applied perpendicular to the current direction, a voltage is generated in a third direction perpendicular to the other two by the Hall effect. This voltage and the current it can produce in an external circuit have many practical applications.

One simple application[1] is in measuring magnetic fields in the air gaps of such devices as relays. Since the crystals can be cut very thin, they can be put in spaces as small as 0.15 mm and can be used to measure the variation in magnetic field with time for many electromechanical devices. Figure 1 shows a photograph of a Hall effect probe, and Fig. 2 shows the relay with a germanium unit mounted in the air gap. The sensitivity of n type germanium is sufficient to allow the unit to be connected directly to cathode-ray oscillographs such as the Dumont 304-A using only the amplifiers that are contained in such oscillographs. For example, a unit 6.0 mm in the direction of the current, 0.15 mm in the direction of the magnetic field, and 1.25 mm in the voltage pick-up direction, with a resistivity of 10 ohm-centimeters for the germanium, generates a voltage given by

$$V/l_z = R_H I H, \tag{1}$$

where R_H is the Hall effect constant, l_z the thickness along the z or voltage pick-up axis, I the current density of the current flowing in the x direction, and H the value of the magnetic field in the y direction. For n type germanium, the measured value of the Hall coefficient divided by the resistivity of the material is

$$R_H/\rho_0 = 3.58\times10^{-5}, \tag{2}$$

when V is expressed in volts and I in amperes per sq cm. This is a constant of the material as long as ρ_0 is greater than 1 ohm-cm and less than 20 ohm-cm. Hence for a resistivity of 10 ohm-centimeters and a current of one milliampere flowing through the crystal, a voltage is generated equal to

$$V = \frac{l_z\times3.58\times10^{-5}\times10\times0.001\times H}{l_y l_z} = 2.39\times10^{-5}H. \tag{3}$$

[1] G. L. Pearson, Rev. Sci. Instr. **19**, 263–265 (1948).

For a field of 1000 gauss, the voltage generated is 0.024 volt which is enough to give a displacement of 0.8 inch on a Dumont 304-A type oscilloscope.

When the crystal is to be used to pick-up time varying fields, the leads from the crystal, as shown by Fig. 1, have to be brought out so that they do not constitute a magnetic loop. This has been done by making the outer frame of metal and soldering the back surface to the frame. Then the voltages for the two loops are opposite and cancel. Since the Hall effect varies somewhat with temperature, the device is usually calibrated by putting it into a permanent magnet gap of known flux and adjusting the oscilloscope amplifier to give the desired sensitivity. Figure 3 is a photograph of several oscilloscope traces connected with the operation of the relay. Figure 3B gives a measure of the magnetic field in the air gap when the armature is prevented from

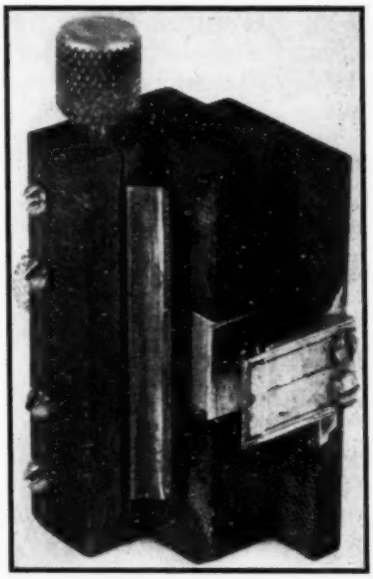


FIG. 1. Photograph of Hall effect probe.

moving. Figure 3E shows the modification caused by the motion of the armature, while Fig. 3F shows the voltage when no current is flowing through the crystal and hence shows that the loop pick-up is quite small compared to the Hall effect voltage. In order to time the motion, Fig. 3D is a 1000-cycle wave put on one of the contacts and shows simultaneously the timing and the time of operation of the contact. Figures 3A and 3C show the current through the relay with the armature clamped and free, respectively.

Another use of the Hall effect is in producing a pure product modulator which is inherently balanced and has no drift characteristic. This use follows from Eq. (1) which shows that the voltage output is directly proportional to the product of the field by the value of the current density I. Hence, if we apply the current of one variable to the magnetic field and the current of

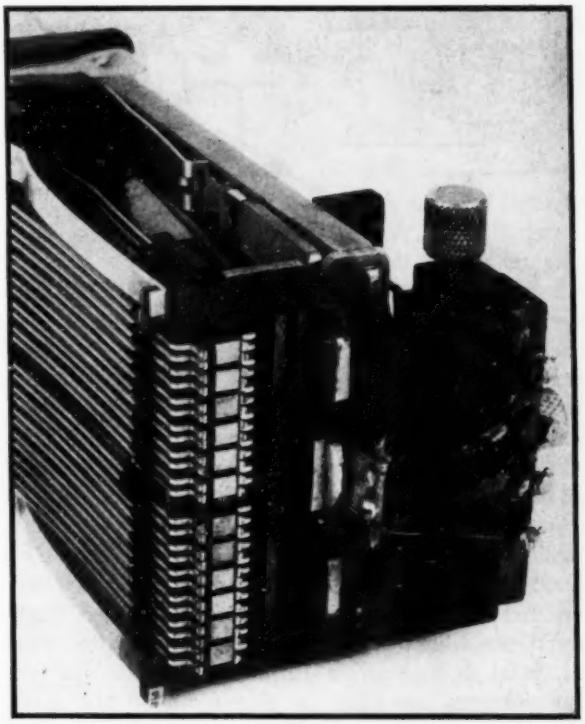


FIG. 2. Use of Hall effect probe in a relay.

the other variable through the crystal, the voltage output is proportional to the product of the two. If the magnet is made of highly laminated material or of a ferrite, the output of the modulator will consist quite exactly of the first-order side bands of the signal applied to the crystal with very little higher order modulation. To get a high linearity, particular crystal orientations

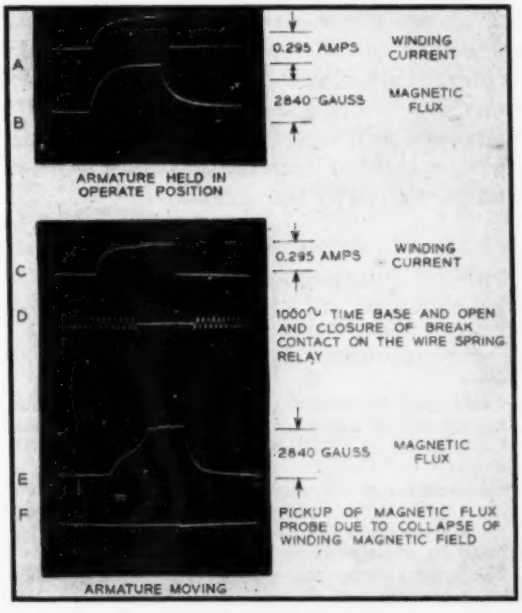


FIG. 3. Photographs of several oscilloscope traces connected with the relay operation.

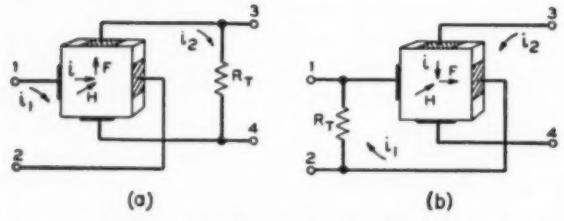


FIG. 4. Figure showing the nonreciprocal nature of Hall effect transmission.

have to be used as discussed in the next section. With a field as high as 10 000 gauss, a modulation of 5 percent can be obtained with the device. While this is low compared to other type modulators, the inherent balance present in the germanium crystal more than makes up for this defect for a number of applications. For example, a nondrift dc amplifier can be obtained by putting a carrier frequency of 1000 cycles on the magnetic circuit and applying the signal to the 1–2 plates of the crystal. When the modulated carrier is amplified in an ac amplifier and the amplified modulated signal demodulated in conventional demodulators, gains as high as 70 db have been obtained with error voltages less than 20 microvolts in the input. With this type of amplifier the zero set does not drift with time or temperature.

The third and probably most interesting application is as a nonreciprocal[2] electrical element which has been called a "gyrator."[3] When the force-current-magnetic field rule is applied, it is readily seen from Fig. 4A that current flowing in a clockwise direction in the 1, 2 circuit will cause a clockwise flow of current in the 3, 4 circuit. On the other hand, as shown by Fig. 4B, applying the same rule, a counter-clockwise flow of current in the 3, 4 circuit causes a clockwise flow of current in the 1, 2 circuit. Hence a zero phase shift occurs for a current flow in one direction and a 180° phase shift occurs for the current flow in the opposite direction and the reciprocity theorem is violated. With the currents and voltages directed as shown by Fig. 5, the relation between the currents and voltages in a symmetrical "gyrator" are given by the equations

$$\begin{aligned} E_1 &= R_{11} i_1 + R_{12} i_2, \\ E_2 &= R_{21} i_1 + R_{11} i_2, \end{aligned} \tag{4}$$

where $R_{21} = -R_{12}$. Owing to the fact that the electrodes cover approximately half of each side, and that for high flux densities magneto-resistance effects enter, it is not easy to express the values of R_{11} and R_{12} in terms of the resistivities and Hall effect constants.[4] For the gyrator with permanent magnet giving a flux density of 17 500 lines per square centimeter, whose drawing is shown by Fig. 6, the measured values obtained were

$$R_{11} = R_{22} = 340 \text{ ohms}; \quad -R_{12} = R_{21} = 78 \text{ ohms}. \tag{5}$$

When terminated in its optimum termination of

$$R_T = (R_{11}^2 + R_{12}^2)^{\frac{1}{2}} = 348 \text{ ohms} \tag{6}$$

on each end, the insertion loss for the gyrator is given by the equation

$$\text{I.L.} = \frac{R_{21}}{R_{11} + (R_{11}^2 + R_{12}^2)^{\frac{1}{2}}} = \frac{78}{688} = 0.1132, \tag{7}$$

corresponding to a loss of 19 db in each direction.

The gyrator has the interesting property that, when it is shunted by parallel resistors R_S as shown by Fig. 7, and these resistors are adjusted to a value

$$2R_S = (R_{11}^2 + R_{12}^2)/R_{21}, \tag{8}$$

the transmission in one direction is improved slightly,

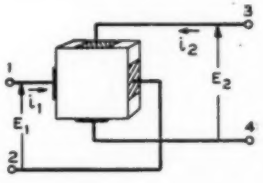


FIG. 5. Direction of currents and voltages used.

while the transmission in the other direction is zero. This follows from the fact that the transmission through the gyrator in the reverse direction is out of phase with that through the resistance network and cancellation takes place. When the network equations (4) are combined with those for the shunting resistances, it is readily shown that the equations for the combination become

$$\begin{aligned} E_1 &= R_{11}' i_1 + R_{12}' i_2, \\ E_2 &= R_{21}' i_1 + R_{11}' i_2, \end{aligned} \tag{9}$$

where

$$R_{12}' = 0; \quad R_{11}' = \frac{(R_{11}^2 + R_{21}^2)^2}{R_{11}^3 + R_{11}^2 R_{21} + 2R_{21}^3};$$

$$R_{21}' = \left(\frac{2R_{21}}{R_{11} + R_{21}}\right)\left(\frac{(R_{11}^2 + R_{21}^2)^2}{R_{11}^3 + R_{11}^2 R_{21} + 2R_{21}^3}\right).$$

[2] The fact that transmission through a crystal by the Hall effect constituted a nonreciprocal electrical system was first pointed out by E. M. McMillan, J. Acoust. Soc. Am. 19, 922 (1947).

[3] The advantages of nonreciprocal circuit elements have been pointed out by B. D. N. Tellegen [Phillips Research Rept. 3, 81 (1948)] who gave them the name of gyrator to symbolize their resemblance to a nonreciprocal mechanical element, the gyroscope. The present name "resistance-gyrator" designates that the network is a nonreciprocal element of the resistive type. Methods for realizing a gyrator at microwave frequencies are discussed by C. L. Hogan, Bell System Tech. J. 31, 1–32 (1952). This method uses the Faraday effect.

[4] An exact solution of the Hall effect voltages and currents for any positions of the electrodes has been obtained by one of the writers, R. F. Wick, and will be discussed in a subsequent paper. This solution shows that even for an infinite magnetic field and a high drift velocity, R_{21} cannot exceed R_{11} and hence an ideal "resistance-gyrator" cannot have a loss of less than 7.6 db. In this respect, the "resistance-gyrator" is deficient compared to a gyrator built from reactive elements which ideally can have a zero loss.

Hence the insertion loss for the shunted "gyrator" is

$$\text{I.L.}=\frac{R_{21}}{R_{11}+R_{21}} \quad \text{and} \quad \text{I.L.}=\frac{R_{12}}{R_{11}+R_{12}}$$

for the two directions.

For the values given in Eq. (5), the insertion loss becomes 14.6 db in one direction and infinity in the other.

Experimentally, a difference in transmission of 60 db for the two directions is easily maintained, and this balance holds over very wide frequency ranges since the Hall effect depends on collision whose mean time is about 10^{-14} second. The limiting frequency of the resistance-gyrator is determined by the condition that dielectric flow is equal to resistive flow. With a dielectric constant of 19 and a resistivity of 10 ohm-centimeters, this frequency limit is about 10^{10} cycles

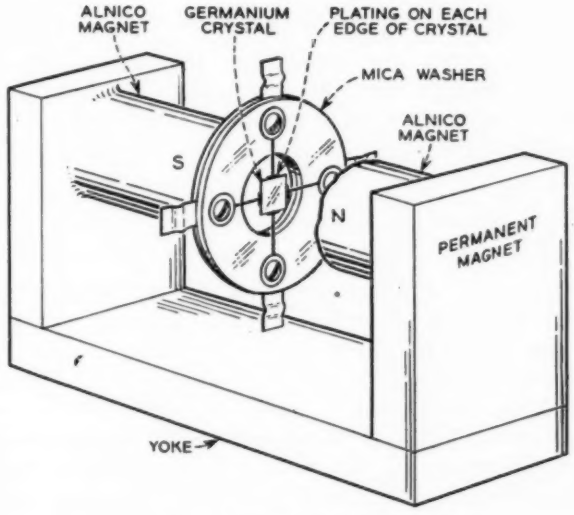


FIG. 6. Drawing of "gyrator."

Hence, it appears that the "gyrator" may be an important circuit element.

II. PHENOMENOLOGICAL DERIVATION OF MAGNETO-RESISTANCE AND HALL EFFECT TERMS

In all the devices described above, it is desirable to have a Hall effect voltage which is accurately proportional to the magnetic field up to high field values. Furthermore, it does not seem to have been pointed out that for most crystal orientations, a voltage can be obtained in a direction perpendicular to the field and applied current by the magneto-resistance effect, and this voltage, which varies as the square of the applied field, is of the same order as the Hall effect voltage when the applied fields are large. This voltage, which is the largest source of distortion in Hall effect devices can be removed by choosing the correct set of crystal orientations.

The next most prominent distortion is produced by

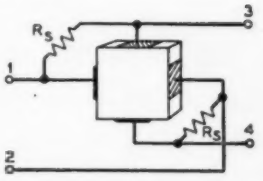


FIG. 7. Gyrator with paralleling resistances for zero transmission in one direction and finite transmission in the other.

the variation of the Hall constant with field. In certain orientations, the Hall constant may drop 10 percent from a low field value to a field of 10 000 gauss. However, measurements reported here show when the field is applied along one of the crystal axes, the Hall constant does not change by as much as 1 percent up to a field of 10 000 gauss and not more than 2 percent up to the fields of 20 000 gauss.

In order to determine what measurements to take to characterize the magneto-resistance and Hall effect corrections, a phenomenological derivation has been given for these effects. A derivation for the first-order magneto-resistive effects has been given by Seitz,[5] who writes the equations in the form

$$\begin{aligned} I_1&=\sigma_0 E_1+\alpha[E_2H_3-E_3H_2]+\beta E_1H^2 \\ &\quad +\gamma H_1[\mathbf{E}\cdot\mathbf{H}]+\delta E_1H_1^2, \\ I_2&=\sigma_0 E_2+\alpha[E_3H_1-E_1H_3]+\beta E_2H^2 \\ &\quad +\gamma H_2[\mathbf{E}\cdot\mathbf{H}]+\delta E_2H_2^2, \\ I_3&=\sigma_0 E_3+\alpha[E_1H_2-E_2H_1]+\beta E_3H^2 \\ &\quad +\gamma H_3[\mathbf{E}\cdot\mathbf{H}]+\delta E_3H_3^2, \end{aligned} \tag{10}$$

where $H^2=H_1^2+H_2^2+H_3^2$, σ is the conductivity, α is related to the Hall coefficient, and β, γ, and δ are three constants related to the magneto-resistance effect.

In order to more easily transform these equations to axes not located along the crystallographic axes, it is desirable to express these equations in tensor form. The three equations of (10) can be written in the tensor form

$$I_i=\sigma_{ij}E_j+\epsilon_{ijk}E_i\alpha_{km}H_m+\beta_{ijkl}E_jH_kH_l, \tag{11}$$

where σ_{ij} and α_{im} are second-rank tensors, β_{ijkl} is a fourth-rank tensor, and ϵ_{ijk} is a third-rank tensor equal to zero when $i=j$ or k or $j=k$, but equal to 1 or -1 when all three numbers are different. If the numbers are in rotation, i.e., 1, 2, 3; 2, 3, 1; 3, 1, 2, the value is $+1$ while if they are out of rotation, the value is -1. Since germanium is a cubic crystal of class 0_h having a center of symmetry, the second rank tensors σ_{ij} and α_{ij} have the terms shown by the first matrix of (12), while the fourth-rank tensor has the terms shown by the second matrix of (12):

$$\begin{matrix} & & \sigma_{11} & 0 & 0 & & \\ & & 0 & \sigma_{11} & 0 & & \\ & & 0 & 0 & \sigma_{11} & & \end{matrix}$$

$$\begin{matrix} \beta_{1111} & \beta_{1122} & \beta_{1122} & 0 & 0 & 0 \\ \beta_{1122} & \beta_{1111} & \beta_{1122} & 0 & 0 & 0 \\ \beta_{1122} & \beta_{1122} & \beta_{1111} & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta_{1212} & 0 & 0 \\ 0 & 0 & 0 & 0 & \beta_{1212} & 0 \\ 0 & 0 & 0 & 0 & 0 & \beta_{1212}. \end{matrix} \tag{12}$$

[5] F. Seitz, Phys. Rev. **9**, 32–35 (1950).

If we expand the tensor Eqs. (11) using the terms of Eq. (12), the two Eqs. (10) and (11) agree if

$$\beta_{1122}=\beta;\quad 2\beta_{1212}=\gamma;\quad (\beta_{1111}-\beta_{1122}-2\beta_{1212})=\delta. \quad (13)$$

Seitz[5] has also suggested that corrections to the Hall effect can be taken account of by adding terms of the form

$$\epsilon'[E_2H_3{}^3-E_3H_2{}^3]+\lambda'H_1{}^2[E_2H_3-E_3H_2]+\eta'H_2H_3[E_2H_2-E_3H_3] \quad (14)$$

to the first of Eqs. (10) with corresponding terms to the other two equations. However, it can be shown that this form of correction terms does not satisfy the equations derived from the thermodynamics of irreversible processes.

The form of the Hall effect equations that lends itself most easily to experimental comparison is the one which expresses the electric fields in terms of the magnetic fields and the current densities. By employing the thermodynamics of irreversible processes and Onsager's principle of microscope reversibility (in time), Casimer[6] first showed that the magneto-resistance and Hall effects could be expressed by the tensor equation

$$E_i=\rho_{ij}(H)I_j+\epsilon_{ijk}I_jR_k(H), \quad (15)$$

where ρ_{ij} is an even function of the magnetic field H and R_k an odd function. As a consequence, we can develop these functions in a MacLaurin series about the point of zero magnetic field, obtaining

$$\begin{aligned}\rho_{ij}(H)&=\rho_{ij}+\frac{1}{2!}\frac{\partial^2\rho_{ij}}{\partial H_k\partial H_l}H_kH_l\\&\quad+\frac{1}{4!}\frac{\partial^4\rho_{ij}}{\partial H_k\partial H_l\partial H_m\partial H_n}H_mH_nH_oH_p\cdots,\\R_k(H)&=\frac{\partial R_k}{\partial H_m}H_m+\frac{1}{3!}\frac{\partial R_k}{\partial H_m\partial H_n\partial H_o}H_mH_nH_o\\&\quad+\frac{1}{5!}\frac{\partial R_k}{\partial H_m\partial H_n\partial H_o\partial H_p\partial H_q}H_mH_nH_oH_pH_q+\cdots.\end{aligned} \quad (16)$$

Designating the partial derivatives by the tensor letters

$$\frac{1}{2!}\frac{\partial^2\rho_{ij}}{\partial H_k\partial H_l}=\alpha_{ijkl};\quad \frac{1}{4!}\frac{\partial^4\rho_{ij}}{\partial H_k\partial H_l\partial H_m\partial H_n}=\zeta_{ijklmn}$$

$$\frac{\partial R_k}{\partial H_m}=R_{km};\quad \frac{1}{3!}\frac{\partial R_k}{\partial H_m\partial H_n\partial H_o}=\gamma_{kmno};$$

$$\frac{1}{5!}\frac{\partial R_k}{\partial H_m\partial H_n\partial H_o\partial H_p\partial H_q}=\lambda_{kmnopq},$$

Eq. (15) can be written in the form

$$\begin{aligned}E_i=&[\rho_{ij}+\alpha_{ijkl}H_kH_l+\zeta_{ijklmn}H_kH_lH_mH_n+\cdots]I_j\\&+\epsilon_{ijk}I_j[R_{km}H_m+\gamma_{kmno}H_mH_nH_o\\&+\lambda_{kmnopq}H_mH_nH_oH_pH_q+\cdots].\end{aligned} \quad (17)$$

When one forms the product $E_1I_1+E_2I_2+E_3I_3$ for these expressions, all terms involving the Hall constants and their correction terms vanish and hence the thermodynamic reversibility criterion is satisfied.

Since germanium is a crystal satisfying the symmetry 0_h (or $[(4/m)\bar{3}(2/m)]$ for the Hermann-Mauguin symbols), the independent constants for the second and fourth-rank tensors ρ_{ij}, R_{lm}, α_{ijkl} are the same as those given by Eq. (12). For the fourth-rank tensor γ_{mnop}, since all four subscripts can be interchanged rather than just the first two or last two, there is one additional relation in (12), namely,

$$\gamma_{1122}=\gamma_{1212}, \quad (18)$$

and there are only two independent constants γ_{1111} and $\gamma_{1122}=\gamma_{1212}$. In order to cut down the number of subscripts, it has become customary to replace reversible pairs of subscripts by single numbers as shown by Eq. (19):

$$11=1;\quad 22=2;\quad 33=3;\quad 23=32=4;\quad 13=31=5;\quad 12=21=6. \quad (19)$$

In a previous paper[7] appendix I, it has been shown that there are 60 nonvanishing terms in a sixth-rank tensor for the symmetry 0_h when the 1–2, 3–4 and 5–6 subscripts can be interchanged. There are many relations between these 60 components and it was shown that, if the first two components of the tensors can be interchanged, i.e., $R_{abc}=R_{bac}$, there are nine independent constants. For the present case we are interested in the terms when

$$R_{abc}=R_{acb},$$

and for this case it is readily shown that the independent terms are

Term	Relation between terms
A	$R_{111}=R_{222}=R_{333}$,
B	$R_{112}=R_{113}=R_{221}=R_{223}=R_{331}=R_{332}=R_{121}=R_{131}=R_{212}=R_{232}=R_{313}=R_{323}$,
C	$R_{122}=R_{133}=R_{211}=R_{233}=R_{311}=R_{322}$,
D	$R_{123}=R_{132}=R_{213}=R_{231}=R_{312}=R_{321}$,
E	$R_{441}=R_{552}=R_{663}=R_{414}=R_{525}=R_{636}$,
F	$R_{442}=R_{443}=R_{551}=R_{553}=R_{661}=R_{662}=R_{424}=R_{434}=R_{515}=R_{535}=R_{616}=R_{626}$,
G	$R_{144}=R_{255}=R_{366}$,
H	$R_{155}=R_{166}=R_{244}=R_{266}=R_{344}=R_{355}$,
I	$R_{456}=R_{465}=R_{546}=R_{564}=R_{645}=R_{654}$.

(20)

[6] H. B. G. Casimer, Revs. Modern Phys. **17**, 343–350 (1945). See also S. R. DeGroot, *Thermodynamics of Irreversible Processes* (Interscience Publishing Company, New York, 1951).

[7] W. P. Mason, Phys. Rev. **82**, 715–723 (1951). S. Bhagavantam [Acta Cryst. **2**, 21 (1949)] had previously shown that there were 9 constants for this type of symmetry but did not give the values.

For the ζ_{ijklmn} tensor of Eq. (17) any of the indices k, l, m, or n can be interchanged since they refer to the same variable, the magnetic field. By using the six index forms of the relation of (20), it is easily seen that there are three relations

$$B=H;\quad D=G;\quad E=I; \tag{21}$$

and taking the first terms of each sequence, the independent terms are

$$\zeta_{111},\ \zeta_{112},\ \zeta_{122},\ \zeta_{123},\ \zeta_{441},\ \zeta_{442}. \tag{22}$$

For the sixth-rank tensor λ_{kmnopq} of Eq. (17), all of the last five indices can be interchanged since they refer to the magnetic field, and since the first and second can be interchanged, it is then possible to interchange all six indices. This results in three more relations

$$B=C=F=H;\quad D=E=G=I, \tag{23}$$

and there are three independent terms

$$\lambda_{111},\ \lambda_{112},\ \text{and}\ \lambda_{123}. \tag{24}$$

Introducing these terms into the tensor Eq. (17), the components of the electric field along x can be written in the form, using the reduced indices shown by Eq. (25). The other terms can be obtained by inserting the corresponding matrix terms for the other two directions:

$$\begin{aligned}E_1=&\rho_1 I_1+\alpha_{11}I_1H_1^2+\alpha_{12}I_1(H_2^2+H_3^2)\\&+2\alpha_{66}H_1(I_2H_2+I_3H_3)+\zeta_{111}I_1H_1^4\\&+6\zeta_{112}I_1H_1^2(H_2^2+H_3^2)+\zeta_{122}I_1(H_2^4+H_3^4)\\&+6\zeta_{123}I_1H_2^2H_3^2+12\zeta_{441}H_1H_2H_3(I_2H_3+I_3H_2)\\&+4\zeta_{442}H_1[I_3H_3(H_1^2+H_3^2)+I_2H_2(H_1^2+H_2^2)]\\&+[R_1+3\gamma_{12}H^2+5\lambda_{112}H^4](I_2H_3-I_3H_2)\\&+(\gamma_{11}-3\gamma_{12})(I_2H_3^3-I_3H_2^3)+(\lambda_{111}-5\lambda_{112})\\&\times(I_2H_3^5-I_3H_2^5)+(30\lambda_{123}-10\lambda_{112})\\&\times[I_2H_3H_1^2H_2^2-I_3H_2H_1^2H_3^2]+\cdots.\end{aligned} \tag{25}$$

To be of use in connection with the measurements performed in Sec. III, these equations have to be transformed to other sets of axes which conform to the orientations of the measured crystals. These crystals were made in the form of oriented cylinders, and by rotating them around their axes, a series of measurements were obtained from which the Hall voltage was obtained. In these measurements one applies the current along the x_1' axis, measures the Hall voltage along the x_2' axis, and applies a magnetic field along the x_3' axis. To determine the relations existing, we have to obtain expressions for E_1' and E_2'. To determine these fields, we make use of the transformation equation for a first-rank tensor

$$E_i'=\partial x_i'E_i/\partial x_i, \tag{26}$$

where $\partial x_i'/\partial x_i$ are the direction cosines that are related for the rotated and fixed axes by the matrix

$$\begin{array}{cccc} & x_1 & x_2 & x_3 \\ x_1' & \dfrac{\partial x_1'}{\partial x_1}=l_1; & \dfrac{\partial x_1'}{\partial x_2}=m_1; & \dfrac{\partial x_1'}{\partial x_3}=n_1 \\ x_2' & \dfrac{\partial x_2'}{\partial x_1}=l_2; & \dfrac{\partial x_2'}{\partial x_2}=m_2; & \dfrac{\partial x_2'}{\partial x_3}=n_2 \\ x_3' & \dfrac{\partial x_3'}{\partial x_1}=l_3; & \dfrac{\partial x_3'}{\partial x_2}=m_3; & \dfrac{\partial x_3'}{\partial x_3}=n_3. \end{array} \tag{27}$$

For E_1' and E_2' the equations are

$$E_1'=l_1E_1+m_1E_2+n_1E_3;\ E_2'=l_2E_1+m_2E_2+n_2E_3. \tag{28}$$

We also have to transform the current and field components into the new axis. Since the only current flowing is that along x_1', and the only magnetic field is that along x_3', we have

$$\begin{aligned}&I_1=l_1I_1';\quad I_2=m_1I_1';\quad I_3=n_1I_1';\\&H_1=l_3H_3';\quad H_2=m_3H_3';\quad H_3=n_3H_3'.\end{aligned} \tag{29}$$

Introducing these values into the first of Eqs. (25), and reducing and collecting terms by making use of the 21 relations between the direction cosines of the types given by Eq. (30), the expression for E_1' reduces to that given in Eq. (31):

$$\begin{aligned}&l_1^2+m_1^2+n_1^2=1;\quad l_1l_2+m_1m_2+n_1n_2=0;\\&\qquad\qquad l_1=m_2n_3-n_2m_3.\end{aligned} \tag{30}$$

Inserting these direction cosine relations in the expression for E_1', we find

$$\begin{aligned}E_1'=&\rho_1I_1'+\alpha_{12}I_1'H_3'^2+\zeta_{122}I_1'H_3'^4\\&+(l_1^2l_3^2+m_1^2m_3^2+n_1^2n_3^2)[(\alpha_{11}-\alpha_{12}-2\alpha_{66})\\&\times I_1'H_3'^2+(6\zeta_{112}-4\zeta_{442})I_1'H_3'^4]\\&+(\zeta_{111}-6\zeta_{112}-\zeta_{122})(l_1^2l_3^4+m_1^2m_3^4+n_1^2n_3^4)\\&\times I_1'H_3'^4+6\zeta_{123}(l_1^2m_3^2n_3^2+m_1^2l_3^2n_3^2+n_1^2l_3^2m_3^2)\\&\times I_1'H_3'^4+(24\zeta_{441}-8\zeta_{442})(l_1l_3m_1m_3n_3^2\\&\qquad+l_1l_3n_1n_3m_3^2+m_1m_3n_1n_3l_3^2)I_1'H_3'^4.\end{aligned} \tag{31}$$

Hence the only terms that affect the resistivity as long as I_2' is zero are the initial resistivity ρ_1 and the first- and second-order magneto-resistance terms.

Similarly, if we calculate the Hall voltage gradient E_2' from the formula $E_2'=l_2E_1+m_2E_2+n_2E_3$, we find on collecting terms

$$\begin{aligned}E_2'=&[(\alpha_{11}-\alpha_{12}-2\alpha_{66})I_1'H_3'^2+(6\zeta_{112}-4\zeta_{442})\\&\times I_1'H_3'^4](l_1l_2l_3^2+m_1m_2m_3^2+n_1n_2n_3^2)\\&+(\zeta_{111}-6\zeta_{112}-\zeta_{122})I_1'H_3'^4(l_1l_2l_3^4\\&+m_1m_2m_3^4+n_1n_2n_3^4)+6\zeta_{123}I_1'H_3'^4\\&\times(l_1l_2m_3^2n_3^2+m_1m_2l_3^2n_3^2+n_1n_2l_3^2m_3^2)\\&+(12\zeta_{441}-4\zeta_{442})I_1'H_3'^4l_3m_3n_3\\&\times[l_2(m_1n_3+n_1m_3)+m_2(l_1n_3+n_1l_3)\\&+n_2(l_1m_3+m_1l_3)]-[R_1+3\gamma_{12}H_3'^2\\&+5\lambda_{112}H_3'^4]I_1'H_3'-(\gamma_{11}-3\gamma_{12})I_1'H_3'^3\\&\times(l_3^4+m_3^4+n_3^4)-(\lambda_{111}-5\lambda_{112})\\&I_1'H_3'^5(l_3^6+m_3^6+n_3^6)-[90\lambda_{123}-30\lambda_{112}]\\&\qquad\times I_1'H_3'^5l_3^2m_3^2n_3^2.\end{aligned} \tag{32}$$

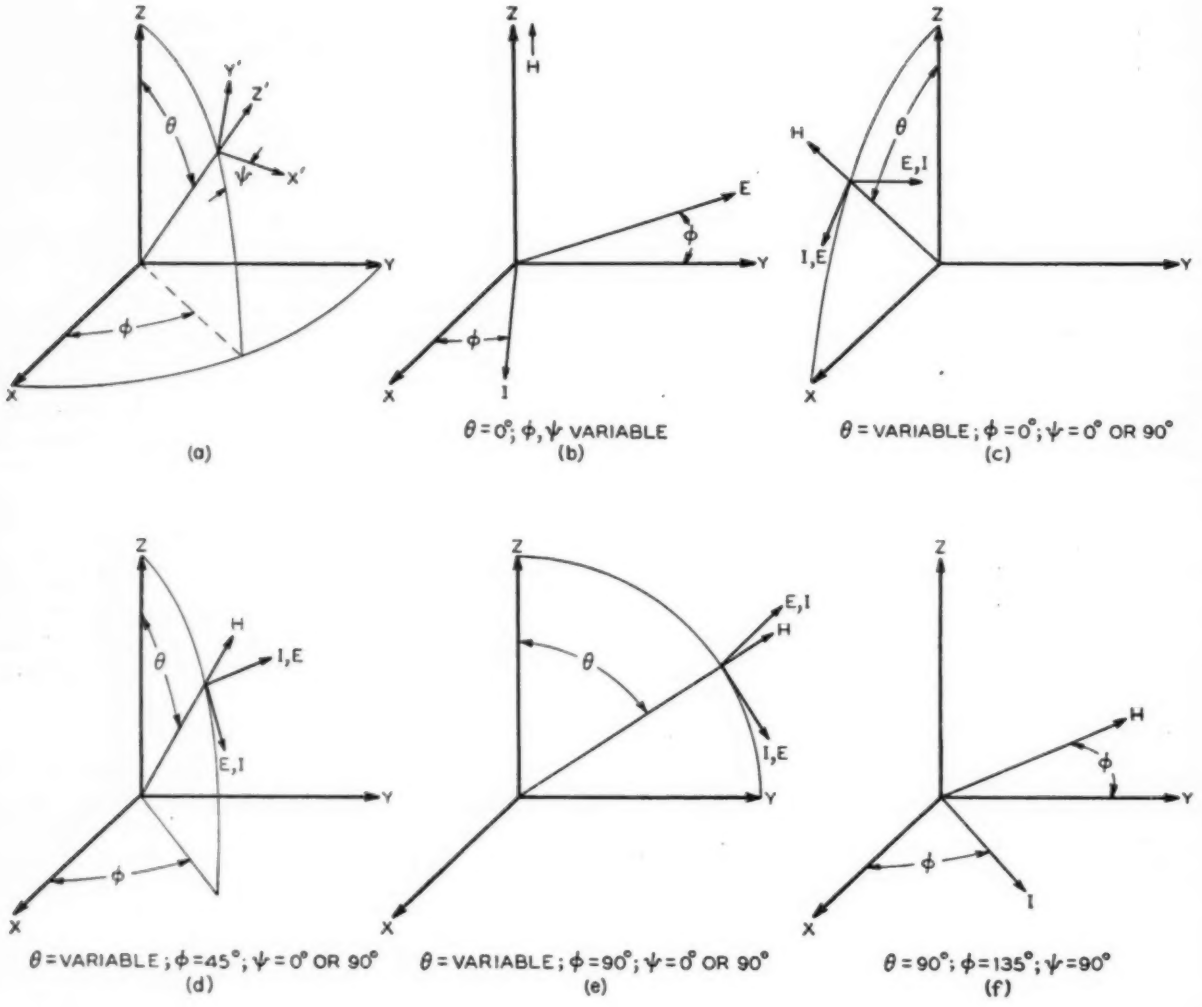


FIG. 8. Orientations for zero cross magneto-resistance effect.

The first four terms show that there can be a cross voltage E_2' even in the absence of a Hall effect due to the magneto resistance effect and this voltage does not change sign when the magnetic field is reversed. These terms come in due to the anisotropic nature of the material and, as will be shown later, this effect disappears for certain angles of cut with respect to the crystallographic axes. In all the experimental measurements taken, the angles are such that this effect disappears and all the cross voltage is the result of the Hall effect.

One further condition is of interest if we wish to evaluate all of the magneto-resistance coefficients. To evaluate the coefficients α_{11} and ζ_{111}, we have to introduce a field in the same direction as the current. Hence, we calculate E_3' when the current density is I_3' and the magnetic field H_3'. Then

$$I_1=l_3I_3'; \quad I_2=m_3I_3'; \quad I_3=n_3I_3' \tag{33}$$

and

$$E_3'=l_3E_1+m_3E_2+n_3E_3. \tag{34}$$

Introducing these terms and combining,

$$\begin{aligned}E_3'=\rho_1I_3'+[(\alpha_{11}-\alpha_{12}-2\alpha_{66})(l_3^4+m_3^4+n_3^4)\\+(\alpha_{12}+2\alpha_{66})]I_3'H_3'^2+[(\zeta_{111}-6\zeta_{112}-\zeta_{122}-8\zeta_{442})\\\times(l_3^6+m_3^6+n_3^6)+(6\zeta_{112}+\zeta_{122}+8\zeta_{442})\\\times(l_3^4+m_3^4+n_3^4)+(18\zeta_{123}+36\zeta_{441})\\\times l_3^2m_3^2n_3^2]I_1'H_3'^4.\end{aligned} \tag{35}$$

Hence if we measure along any of the crystallographic axes,

$$E_3'=\rho_3I_3'+\alpha_{11}I_3'H_3'^2+\zeta_{111}I_3'H_3'^4 \tag{36}$$

and the constants α_{11} and ζ_{111} can be evaluated.

In determining the actual values, it is desirable to introduce the three angles of Fig. 8(a) in place of the direction cosines. In terms of these angles it is well known that the direction cosines are given by the

TABLE I. Relation between crystal directions and resistivity equations.

Direction of I	Direction of H	θ	φ	ψ	Equations for resistivity
100	100	90°	0°	0°	$\rho=\rho_0+\alpha_{11}H_3'^2+\zeta_{111}H_3'^4$
100	010	0°	0°	90°	$\rho=\rho_0+\alpha_{12}H_3'^2+\zeta_{112}H_3'^4$
110	001	0°	45°	0°	$\rho=\rho_0+\alpha_{12}H_3'^2+\zeta_{122}H_3'^4$
110	110	90°	45°	0°	$\rho=\rho_0+\left(\frac{\alpha_{11}+\alpha_{12}+2\alpha_{66}}{2}\right)H_3'^2+\left(\frac{\zeta_{111}+6\zeta_{112}+\zeta_{122}+8\zeta_{442}}{4}\right)H_3'^4$
110	1$\bar{1}$0	90°	135°	−90°	$\rho=\rho_0+\left(\frac{\alpha_{11}+\alpha_{12}-2\alpha_{66}}{2}\right)H_3'^2+\left(\frac{\zeta_{111}+6\zeta_{112}+3\zeta_{122}-8\zeta_{442}}{4}\right)H_3'^4$

equations

$$l_1=\cos\theta\cos\varphi\cos\psi-\sin\varphi\sin\psi;\quad l_2=-\cos\theta\cos\varphi\sin\psi-\sin\varphi\cos\psi;\quad l_3=\cos\varphi\sin\theta,$$
$$m_1=\cos\theta\sin\varphi\cos\psi+\cos\varphi\sin\psi;\quad m_2=\cos\varphi\cos\psi-\sin\varphi\sin\psi\cos\theta;\quad m_3=\sin\varphi\sin\theta,$$
$$n_1=-\sin\theta\cos\psi;\quad n_2=\sin\theta\sin\psi;\quad n_3=\cos\theta. \tag{37}$$

With these values, the magneto-resistance values and the Hall voltage correction terms can be evaluated for any angular system of axes.

One result of these equations immediately obvious is that in order to eliminate the cross voltage resulting from the magneto-resistance effect a particular set of orientations has to be chosen. These orientations have to be chosen so that the first four terms of Eq. (32) are simultaneously zero. As seen from the values given in Eqs. (37) all these equations are simultaneously zero when $\theta=0$ which corresponds to impressing the magnetic field along one of the crystal axes. All four terms vanish when $\psi=0$ or $\pm90°$, $\varphi=0$, 45°, 90°, etc. As shown by Fig. 8(b) to (f), these values can be realized by impressing the magnetic field in planes containing two of the crystallographic axes, or a plane 45° between them and sending the currents in these planes at right angles to the magnetic fields while measuring the voltages perpendicular to them, or vice versa. These were the conditions used in all of the measurements of the Hall voltage so that the Hall voltages do not have to be corrected for this effect.

III. EVALUATION OF THE CONSTANTS IN THE MAGNETO-RESISTANCE AND THE HALL EFFECT EQUATIONS

The most complete measurements of the magneto-resistance effects are those made by Pearson and Suhl.[8] For small values of the magnetic field, the variations are proportional to the square of the field but for larger fields fourth and sixth power terms have to be added. The second and fourth power terms can represent the measured values well up to magnetic fields of 15 000 gausses.

Five orientations were measured, two with their lengths along the 100 direction and with the field applied, respectively, along the 100 direction and the 010 direction, and three with the length along the 110 direction with the field along the 1$\bar{1}$0 direction, the 001 direction, and the 110 direction. These orientations and the corresponding θ, φ, ψ angles are shown by Table I. Inserting the direction cosines, the equations for the magneto-resistance effect can be obtained from Eq. (31) and the formulas are also given in Table I.

Up to 15 000 gauss, Pearson and Suhl's measurements of the resistivity are best fitted by the equations

$$\begin{aligned}
&I,\ H\\
&100\ 100;\ \rho-\rho_0/\rho_0=1.6\times10^{-9}H^2-1.5\times10^{-18}H^4,\\
&100\ 010;\ \rho-\rho_0/\rho_0=0.7\times10^{-9}H^2-0.6\times10^{-18}H^4,\\
&110\ 1\bar{1}0;\ \rho-\rho_0/\rho_0=1.5\times10^{-9}H^2-2.3\times10^{-18}H^4,\\
&110\ 001;\ \rho-\rho_0/\rho_0=0.7\times10^{-9}H^2-0.6\times10^{-18}H^4,\\
&110\ 110;\ \rho-\rho_0/\rho_0=0.8\times10^{-9}H^2-1.7\times10^{-18}H^4.
\end{aligned} \tag{38}$$

From these measurements the constants α_{mn}/ρ_0 and ζ_{mno}/ρ_0 can be evaluated. These are constants independent of the resistivity of the material and for n type germanium the values obtained are

$$\alpha_{11}/\rho_0=1.6\times10^{-9};\quad \alpha_{12}/\rho_0=0.7\times10^{-9};\quad \alpha_{66}=-0.35\times10^{-9};$$
$$\zeta_{111}/\rho_0=-1.5\times10^{-18};\quad \zeta_{122}/\rho_0=-0.6\times10^{-18};\quad \zeta_{112}/\rho_0=-0.88\times10^{-18};\quad \zeta_{442}/\rho_0=+0.075\times10^{-18}. \tag{39}$$

These measurements allow the evaluation of all of the first-order magneto-resistance constants and four of the six-second-order constants. To evaluate the other two constants, two other independent orientations would have to be measured.

In order to evaluate the variation of the Hall effect as a function of angle, two cylinders were cut out of a single germanium crystal, one having its axis along a [100] direction, and the other with its axis along a [110] direction. The dimensions of the crystals were, respectively, length = 1.623 cm, diameter 0.292 cm and length = 1.610 cm, diameter = 0.298 cm. The Hall effect voltages were measured by putting probes on the cylinder at right angles to the field and measuring by a potentiometer the voltage generated. To take account of the fact that the specimen was cylindrical, we have from the definition of the Hall effect

$$R_H=\frac{E/D}{IH}=\frac{E/D}{(4i/\pi D^2)H}=\frac{\pi}{4}\left(\frac{ED}{iH}\right), \tag{40}$$

[8] G. L. Pearson and H. Suhl, Phys. Rev. 83, 768–776 (1951).

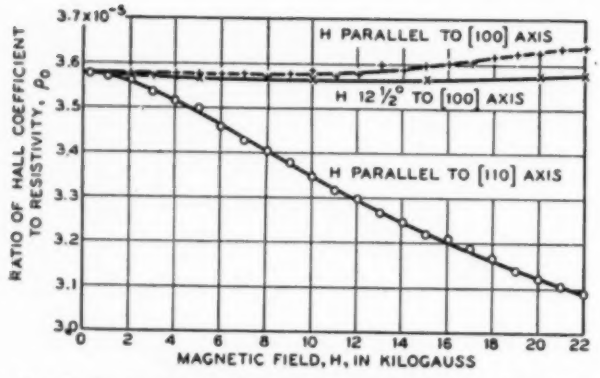

RATIO OF HALL COEFFICIENT TO RESISTIVITY, ρ_0
H PARALLEL TO [100] AXIS
H 12½° TO [100] AXIS
H PARALLEL TO [110] AXIS
MAGNETIC FIELD, H, IN KILOGAUSS

FIG. 9. Ratio of R_H/ρ_0 for a [100] cylinder for two directions of the magnetic field.

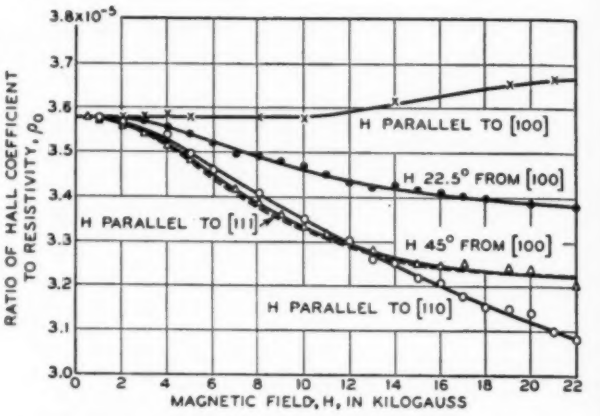

RATIO OF HALL COEFFICIENT TO RESISTIVITY, ρ_0
H PARALLEL TO [100]
H 22.5° FROM [100]
H PARALLEL TO [111]
H 45° FROM [100]
H PARALLEL TO [110]
MAGNETIC FIELD, H, IN KILOGAUSS

FIG. 10. Ratio of R_H/ρ_0 for a [110] cylinder for five directions of the magnetic field.

where D is the diameter, E the measured voltage, I the current density, i the total current, and H the field. The material was 3.18 ohm-centimeter material, and to obtain an absolute constant, the ratio of the Hall voltage to the resistivity is plotted in the measurements made. The magnetic field was put on by a large electromagnet which had essentially a constant field over the length of the cylinder. The field was calibrated by using a rotating coil driven by a synchronous motor and measuring the voltage generated.

Figure 9 shows the ratio of R_H to ρ_0 plotted as a function of the magnetic field for the [100] cylinder. The top curve is for the magnetic field along the [001] direction, and, as can be seen, the Hall constant does not vary appreciably with field out to 10 000 gauss and increases by only 2 percent out to a field of 22 000 gauss. For the field along the [011] direction the Hall coefficient decreases by 6.5 percent out to 10 000 gauss and 14 percent at 22 000 gauss. An intermediate case is also shown for which H is 12.5 degrees from the [100] axis. This angle gives the smallest deviation from a straight line relation for fluxes as high as 22 000 gauss. The top and bottom curves can be represented by the equation valid up to 15 000 gauss,

$$\begin{aligned} &H \text{ along } [001];\ R_H/\rho_0=3.58\times10^{-5}[1+2.4\times10^{-19}H^4],\\ &H \text{ along } [011];\ R_H/\rho_0=3.58\times10^{-5}\\ &\qquad\times[1-7.9\times10^{-10}H^2+1.55\times10^{-18}H^4]. \end{aligned} \tag{41}$$

Figure 10 shows the ratio of R_H to ρ_0 plotted as a function of the magnetic field for the [110] cylinder. Five curves were measured with the field varying between the [001] and [1$\bar{1}$0] directions. These can be fitted by the equations

$$\begin{aligned} &H \text{ along } [001];\ R_H/\rho_0=3.58\times10^{-5}[1+2.4\times10^{-19}H^4],\\ &H \text{ 22.5° from } [001];\ R_H/\rho_0=3.58\times10^{-5}\\ &\qquad\times[1-4.13\times10^{-10}H^2+0.92\times10^{-18}H^4],\\ &H \text{ 45° from } [001];\ R_H/\rho_0=3.58\times10^{-5}\\ &\qquad\times[1-8.8\times10^{-10}H^2+1.96\times10^{-18}H^4],\\ &H \text{ 55° from } [001];\ R_H/\rho_0=3.58\times10^{-5}\\ &\qquad\times[1-9.8\times10^{-10}H^2+2.5\times10^{-18}H^4],\\ &H \text{ along } [1\bar{1}0];\ R_H/\rho_0=3.58\times10^{-5}\\ &\qquad\times[1-7.9\times10^{-10}H^2+1.55\times10^{-18}H^4]. \end{aligned} \tag{42}$$

The volts for low values of H reaches a minimum in the [111] direction which agrees with the phenomenological theory given previously. The equations for the Hall voltage along the [100] and [110] cylinders are identical when the fields are along [001] for both cases and for [110] directions for both cases.

All of these orientations are orientations for which the magneto-resistance effects do not produce any cross voltage. Hence the measured values can be used to evaluate the first- and second-order correction terms

TABLE II. Relation between crystal directions and Hall effect equations.

Direction of I	θ	Values of φ	ψ	Equations for the Hall effect
100	0	0	0	$R_H=R_1+\gamma_{11}H^2+\lambda_{111}H^4$
100	45°	90°	−90°	$R_H=R_1+\frac{(\gamma_{11}+3\gamma_{12})}{2}H^2+\frac{(\lambda_{111}+15\lambda_{112})}{4}H^4$
110	0	−45°	90°	$R_H=R_1+\gamma_{11}H^2+\lambda_{111}H^4$
110	22.5°	−45°	90°	$R_H=R_1+(0.7393\gamma_{11}+0.7821\gamma_{12})H^2+(0.623\lambda_{111}+1.75\lambda_{112}+0.414\lambda_{123})H^4$
110	45°	−45°	90°	$R_H=R_1+\left(\frac{3}{8}\gamma_{11}+\frac{15}{8}\gamma_{12}\right)H^2+\left(\frac{5}{32}\lambda_{111}+\frac{105}{32}\lambda_{112}+\frac{90}{32}\lambda_{123}\right)H^4$
110	55°	−45°	90°	$R_H=R_1+\left(\frac{\gamma_{11}+6\gamma_{12}}{3}\right)H^2+\left(\frac{\lambda_{111}+30\lambda_{112}+30\lambda_{123}}{9}\right)H^4$
110	90°	−45°	90°	$R_H=R_1+\left(\frac{\gamma_{11}+3\gamma_{12}}{2}\right)H^2+\left(\frac{\lambda_{111}+15\lambda_{112}}{4}\right)H^4$

TABLE III. Evaluation of Hall effect constants and comparison with measured values.

Direction of I	Value of θ	φ	ψ	Sum of first order terms Meas	Sum of first order terms Calc	Percent error	Sum of second order terms Meas	Sum of second order terms Calc	Percent error
100	0	0	0	0	0	0	$+0.24\times10^{-18}$	$+0.24\times10^{-18}$	0
100	45°	90°	−90°	-7.9×10^{-10}	-7.5×10^{-10}	5 percent	$+1.55\times10^{-18}$	$+1.56\times10^{-18}$	1 percent
110	0	−45°	90°	0	0	0	$+0.24\times10^{-18}$	$+0.24\times10^{-18}$	0
110	22.5°	−45°	90°	-4.13×10^{-10}	-3.5×10^{-10}	6 percent	$+0.92\times10^{-18}$	$+0.97\times10^{-18}$	5.5 percent
110	45°	−45°	90°	-8.8×10^{-10}	-9.4×10^{-10}	6.5 percent	$+1.96\times10^{-18}$	$+2.19\times10^{-18}$	12 percent
110	55°	−45°	90°	-9.8×10^{-10}	-10.0×10^{-10}	2 percent	$+2.5\times10^{-18}$	$+2.34\times10^{-18}$	6.5 percent
110	90°	−45°	90°	-7.9×10^{-10}	-7.5×10^{-10}	5 percent	$+1.55\times10^{-18}$	$+1.56\times10^{-18}$	1 percent

for the Hall effect. For the angles used in the measurements, the theoretical values are given in Table II.

Evaluating the first- and second-order Hall effect constants so that they agree best with the measured data of Eqs. (41) and (42), we find that

$$\gamma_{11}/R_1=0;\quad \gamma_{12}/R_1=-5.0\times10^{-10};\quad \lambda_{111}/R_1=2.4\times10^{-19};\quad \lambda_{112}/R_1=4.0\times10^{-19};\quad \lambda_{123}/R_1=3.0\times10^{-19}. \tag{43}$$

With these values the agreement with the first- and second-order terms for the various orientation is shown by Table III.

The agreement is good to about 6 percent for the first-order terms and 12 percent for the second-order terms, which is probably within the experimental error.

IV. USEFUL CRYSTAL CUTS FOR PRODUCING A HALL COEFFICIENT INDEPENDENT OF THE FIELD

If any arbitrary orientation is used for the measurement or application of the Hall effect, the largest source of error is the cross voltage produced by the magneto-resistance effect. From Eq. (32) it is seen that the worst magneto-resistance cross voltage will occur when $\theta=90°$, $\varphi=45°$, $\psi=45°$, for which case the voltage is equal to

$$E_2'=I_1'H_3'^2\left[\frac{\alpha_{11}-\alpha_{12}-2\alpha_{66}}{4}\right]+\frac{[\zeta_{111}-\zeta_{122}+6\zeta_{112}-8\zeta_{442}-6\zeta_{123}]}{8}I_1'H_3'^4 -\left[R_1+\left(\frac{\gamma_{11}+3\gamma_{12}}{2}\right)H_3'^2+\left(\frac{\lambda_{111}+15\lambda_{112}}{4}\right)H_3'^4\right]I_1'H_3'. \tag{44}$$

While all the second-order terms have not been evaluated, the first-order magneto-resistance term will produce 11 percent of the voltage generated by the Hall effect for a field of 10 000 gauss. The magneto-resistance cross voltage can be eliminated by using one of the orientations shown by Fig. 8.

The next largest source of error is the corrections to the Hall constant. In the above example, this can produce 7.5 percent error for a field of 10 000 gauss and as much as 25 percent for a field of 20 000 gauss. By cutting a 100 section with its length along another crystallographic axis and the field applied along the third, the square law correction is entirely removed, and the fourth power correction is only about 2 percent out to a field of 20 000 gauss. Hence this is the preferred orientation for flux measuring devices, modulators, and gyrators. As shown by Table II, a similar independence of Hall constant with field is obtained when the current is along the [110] direction and the field is along the [001] direction, and hence this is an equally accepted orientation. For large flux densities, the 12½° orientation of Fig. 9 gives the best all round results.

Screw Dislocations in Thin Rods*

J. D. ESHELBY
Department of Physics, University of Illinois, Urbana, Illinois
(Received June 13, 1952)

In connection with Galt and Herring's observation on thin whiskers of tin, the properties of a screw dislocation in a cylinder are worked out. When all boundary conditions are taken into account, the image force tends to keep the dislocation along the axis. Only when it is displaced about half-way to the surface does the image force tend to pull it out of the rod. Generators of the cylindrical rod become helices when the dislocation is introduced. The dislocation can be ejected from the rod by twisting or bending it suitably.

INTRODUCTION

GALT and Herring[1] have described the mechanical properties of thin whiskers of tin. Recent work on the theory of crystal growth[2] suggests that there may be a dislocation with a substantial screw component running parallel to the axis of the whisker. On the other hand, their enhanced strength might suggest that the whiskers are free from dislocations. It may therefore be useful to work out the properties of a rod containing a screw dislocation as a problem in isotropic elastic theory. The results from the theory of the torsion and bending of beams which we use may be found, for example, in Sokolnikoff's[3] book.

We first of all treat the problem as one in antiplane strain in which the displacement w is everywhere parallel to the z axis and is independent of z. The only nonvanishing stress components are τ_{zx} and τ_{zy}, which may be regarded as the x and y components of a vector $\boldsymbol{\tau}=\mu$ grad w, where μ is the shear modulus. The equilibrium condition is div $\boldsymbol{\tau}=\mu\nabla^2 w=0$.

The displacement due to a screw dislocation through the point $(x=\xi, y=0)$ when the surface of the infinite cylinder $x^2+y^2=R^2$ is stress-free is

$$w=\frac{b}{2\pi}\tan^{-1}\frac{y}{x-\xi}-\frac{b}{2\pi}\tan^{-1}\frac{y}{x-R^2/\xi}. \qquad (1)$$

The stress is also zero on any other circle for which $(\xi, 0)$ and $(R^2/\xi, 0)$ are inverse points. Equation (1), therefore, equally well represents the displacement when the core of the dislocation is excluded by a small cylinder of radius r_0. If $r_0 \ll R$, the center of the small circle is practically at $(\xi, 0)$. The w of (1) is multiple-valued: we can make it single-valued by a cut joining the circles r_0 and R, most conveniently along the x axis (Fig. 1).

The elastic energy per unit length of the cylinder is

$$W=\tfrac{1}{2}\mu\int(\text{grad } w)^2 dxdy$$

$$=\tfrac{1}{2}\mu\int_C w\frac{\partial w}{\partial n}ds$$

taken round the circuit C in Fig. 1. If we let r_0 tend to zero and reject an infinite constant independent of ξ, the result is

$$W=\frac{\mu b^2}{4\pi}\ln(R^2-\xi^2).$$

Figure 2(a) shows that the dislocation is in unstable equilibrium when it is at $\xi=0$; at any other point there is an "image-force,"

$$F=-\frac{\partial W}{\partial \xi}=\frac{\mu b^2}{2\pi}\frac{\xi}{R^2-\xi^2},$$

tending to drive it out of the cylinder.

On any cross section of the cylinder there is a couple

$$M=\mu\int\left(x\frac{\partial w}{\partial y}-y\frac{\partial w}{\partial x}\right)dxdy$$

$$=\mu\int_C w(xm-yl)ds,$$

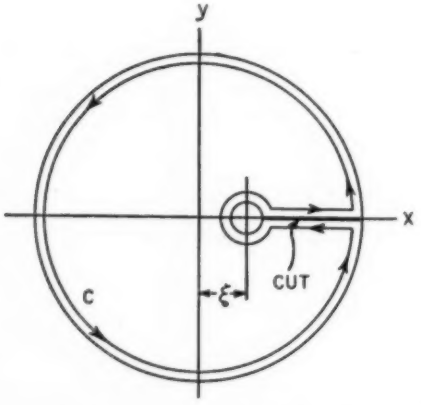


FIG. 1. Circuit used in determining couple and energy.

* Work supported by the U. S. Office of Naval Research.
[1] Conyers Herring and J. K. Galt, Phys. Rev. **85**, 1060 (1952).
[2] See the review article by F. C. Frank, Advances in Physics (Philosophical Magazine Supplement) **1**, 91 (1952).
[3] Ivan S. Sokolnikoff, *Mathematical Theory of Elasticity* (McGraw-Hill Book Company, Inc., New York, 1946).

where (l, m) is the normal to the curve C. When $r_0 \to 0$, this gives

$$M(\xi)=\tfrac{1}{2}\mu b(R^2-\xi^2).$$

We can regard the couple as being produced by the surface tractions on the ends of the cylinder at infinity which are necessary to maintain a state of antiplane strain. If we drop the antiplane strain requirement, we can get rid of M by applying a couple $-M$; this gives a twist in which two cross sections of the cylinder, unit distance apart, undergo a relative rotation

$$\alpha(\xi)=\frac{2M}{\mu\pi R^4}=\frac{b}{\pi R^2}(1-\xi^2/R^2). \tag{2}$$

This reduces the energy by $\frac{1}{2}\alpha M$ per unit length, and W becomes

$$W=\frac{\mu b^2}{4\pi}\left[\ln(R^2-\xi^2)-\frac{(R^2-\xi^2)^2}{R^4}\right]. \tag{3}$$

Figure 2(b) shows that the dislocation is now in stable (strictly, metastable) equilibrium at the center. Only if it is displaced further than

$$\xi_{\max}=(1-2^{-\frac{1}{2}})^{\frac{1}{2}}R=0.54R$$

from the center will the dislocation be pulled out of the cylinder by the image force.

If we now cut out a rod of finite length l from the infinite cylinder, there will be a distribution of surface traction on each end which has zero resultant and zero moment. According to St. Venant's principle, the effects of relaxing these surface tractions will penetrate only to a distance of order R from the ends, so that if $R/l \ll 1$ we may still take (3) as the energy per unit length of the finite cylinder. The effective force on the dislocation is now

$$F=-\frac{\partial W}{\partial \xi}=\frac{\mu b^2}{4\pi R^2}\xi\left[\frac{R^2}{R^2-\xi^2}-2\frac{R^2-\xi^2}{R^2}\right]. \tag{4}$$

It has a maximum and a minimum

$$F=\mp 0.216\frac{\mu b^2}{4\pi R} \quad \text{at} \quad \xi=\pm 0.322R. \tag{5}$$

The force can also be derived directly from the rule that the force on a screw dislocation running parallel to the z axis is

$$F_x=b\tau_{xy}, \quad F_y=-b\tau_{xz}, \tag{6}$$

where the stresses are to be evaluated at the position of the dislocation after rejecting the stresses which the dislocation would produce if the medium were extended to infinity. In the present case τ_{xy} and τ_{xz} are the stresses arising from the second term in (1) and from the twist (2). The energy method we have used leads to less worry over signs.

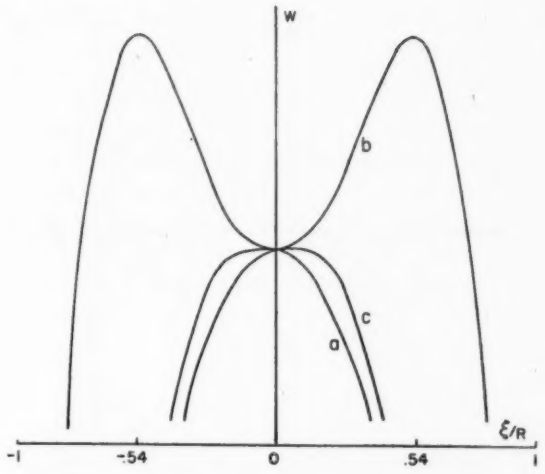


FIG. 2. Energy of a screw dislocation in a cylinder.

BEHAVIOR IN A TWISTED ROD

It is obvious that the point $\xi=0$ again becomes unstable if an external couple sufficient to undo the the twist (2) is applied. We shall show that actually a couple of only half this value is needed. Let a constant external couple M' be applied; to fix ideas, suppose that a weight is hung from a cord wound round one end of the rod. A change in the position of the dislocation will change the strain energy of the rod and raise or lower the weight. To find the equilibrium position we must find the value of ξ which minimizes the sum of the strain energy of the rod and the potential energy of the weight or other external device producing M'. This external energy is $l\alpha(\xi)M'$ in all, or

$$bM'\xi^2/\pi R^4 \tag{7}$$

per unit length of the rod, plus a constant independent of ξ. The strain energy can be found as follows. Before applying M', it is given by Eq. (3). If Hooke's law is valid the body reacts to external forces in the same way whether it has internal stresses or not. Hence, the change in strain energy when M' is applied, being equal to the work done by M', is independent of ξ. In other words, the strain energy of the rod contains no cross term between the externally produced stress and the internal stress arising from the dislocation.[4] Thus, the total energy of the system, but for a constant, is simply the sum of (3) and (7). As M' increases, the maxima of Fig. 2(b) are lowered and shifted nearer the origin. When

$$M'=\tfrac{1}{4}\mu bR^2,$$

the maxima have moved in to the origin, swamping the minimum there and restoring instability [Fig. 2(c)].

The stress-strain curve of the rod in torsion will have the asymmetric form shown in Fig. 3, the horizontal part representing the disappearance of the twist

[4] F. R. N. Nabarro, Advances in Physics **1**, Sec. II 3.5, 269 (1952).

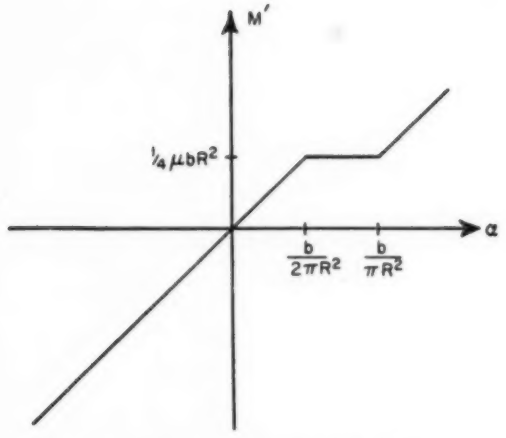


FIG. 3. Ideal torsional stress-strain curve of a whisker containing a dislocation.

when the dislocation leaves the rod. If the sign of b is changed, the curve must be inverted through the origin.

BEHAVIOR IN A BENT ROD

Because of its all-or-nothing effect we cannot use an applied couple to move the dislocation about in a controlled way. This can be done by clamping one end of the rod and deflecting the free end with a transverse load. The deflection of the free end d and the load W are connected by

$$d=\frac{4}{3\pi}\frac{Wl^3}{ER^4},$$

where E is Young's modulus. We shall work in terms of d, since in practice the end of the rod would be deflected by pushing with an adjustable probe rather than by hanging a weight on the end.

If the load acts parallel to the y axis, the force urging the dislocation along the x axis is given by (6) with τ_{zy} equal to the stress in the rod due to bending. The result is

$$F=\tfrac{3}{4}(3+2\nu)\mu b\frac{R^2d}{l^3}\left\{1-\frac{1-2\nu}{3+2\nu}\frac{\xi^2}{R^2}\right\}, \tag{8}$$

where ν is Poisson's ratio. The term in ξ^2 vanishes when $\nu=\frac{1}{2}$, and never reaches $\frac{1}{10}$ of the constant term when $\nu=\frac{1}{3}$.

We may note in passing that for a beam of any section bent in flexure by a load along a principal axis the force on a dislocation is everywhere normal to the lines of shearing stress;[5] these lines are thus contours of an energy surface $W(x,y)$ for which $\mathbf{F}=\text{grad } W$.

To find the position of the dislocation for a given deflection, we have to solve for ξ the equation obtained by setting the sum of (4) and (8) equal to zero. As a sketch of the derivative of curve b of Fig. 2 shows, only the numerically smallest root gives a stable position. For small d

$$\xi=3(3+2\nu)\pi R^4d/bl^3.$$

If (8) exceeds the absolute value of the F of (5), there is no stable position and the dislocation leaves the rod as a result of the bending. The necessary deflection is

$$d=\frac{0.072b}{\pi(3+2\nu)}\left(\frac{l}{R}\right)^3, \tag{9}$$

if we neglect the ξ^2 term in (8).

According to (2) the angle α varies as d (and hence ξ) is altered, so that there is a coupling between the deflection and twist of the rod.

When the dislocation leaves the rod, there will be an additional nonelastic deflection d' which remains when the load is removed. It can be found by equating Wd' to the force (8) integrated from 0 to R. This gives

$$d'=4bl/3\pi R.$$

The plastic bending giving rise to d' will occur at the clamped end of the rod (or at the root of a whisker growing on a metal surface) so that after the dislocation is expelled and the load removed the rod will be inclined at an angle

$$\theta=4b/3\pi R$$

to its original direction.

The type of bending we have just considered (flexion) must be clearly distinguished from, for example, the case in which the rod is bent into a circular arc of radius ρ. In this latter case τ_{zz} and τ_{zy} vanish, and the bending produces no force on the dislocation. However, if, following Mott and Nabarro,[6] we ascribe a line tension μb^2 to the dislocation, there will be an effective force $\mu b^2/\rho$ per unit length perpendicular to the dislocation line. The radius of curvature necessary for this to overcome the maximum image force (4) is

$$\rho=58R,$$

quite a modest degree of curvature for a thin rod. According to elementary beam theory, the maximum strain in a longitudinal filament is then $R/\rho=1.7$ percent. We may also apply this to the flexure problem: the maximum radius of curvature is at the clamped end and equal to $l^2/3d$. The deflection for the dislocation to leave the rod owing to its line tension is thus

$$d=0.0058l^2/R. \tag{10}$$

Whether the dislocation leaves the rod because of this effect or because of the bending stresses [Eq. (9)] depends on the ratio between b/R and R/l.

Attributing a line tension to the dislocation is only a crude approximation to the solution of the difficult problem of calculating the energy of a dislocation line

[5] See, e.g., Ivan S. Sokolnikoff, reference 3, Fig. 43.

[6] Mott and Nabarro, Conference on Strength of Solids, Physical Society (London), 1948, p. 1.

as a function (or rather functional) of its shape and relation to the surface of the body containing it. Moreover, in a strongly curved lattice we must note that strictly the Burgers vector is not constant along the dislocation line, being everywhere parallel to a local lattice vector. Thus, the angle between the Burgers vector and direction of a straight dislocation running through a curved lattice will vary along its length, and so too will the energy per unit length, since it depends on this angle. As a result, the dislocation of lowest energy joining two given points may not be the straight line suggested by the line tension concept, since by deforming it into a curve the inclination of Burgers vector and tangent vector might be given a more favorable value at each point. Hence, it is not clear that the previous calculation is correct even in principle. If it is, the exact numerical results are still doubtful.

POSSIBLE EXPERIMENTS

Galt and Herring[1] found that when bent their "whiskers" behaved nonelastically at a strain of 2 or 3 percent. The conditions of bending probably did not correspond exactly with either of the cases we have discussed. The case of bending into a circular arc gives a maximum longitudinal strain of 1.7 percent. The type of deformation they found (a sharp bend through a large angle) does not agree with the deformation produced by the escape of one dislocation. The rough agreement of the theoretical and experimental critical strains may therefore be an accident, or it may indicate that the movement of the original dislocation triggers off more extensive plastic deformation. According to the calculation of Mackenzie,[7] slip of one atomic plane over another can occur without the aid of dislocations for strains of about 3 percent. Galt and Herring's results are therefore consistent either with there being no dislocation in the whisker or with the presence of a dislocation which plays no part in the deformation.

It is clear that other types of experiment would be more suitable for establishing the presence or absence of a dislocation in the whisker. When the dislocation is along the axis, the twist (2) amounts to about 50° per cm for $b=3\times10^{-8}$ cm and $R=10^{-4}$ cm. It should be possible to detect this by x-ray diffraction. However, it is not merely a question of a slow variation of lattice orientation along the length of the specimen. The relative rotation of the extreme cross sections illuminated by a beam of reasonable width will be quite large. In addition to the twist, there is the displacement $w=(b/2\pi)\tan^{-1}(y/x)$. The diffraction pattern caused by the latter alone has been calculated by Wilson.[8]

It might also be possible to measure the torsional stress-strain curve of a whisker, using a small superimposed tension to prevent the specimen raveling up. For a specimen 1 cm long our simple model suggests that the dislocation should be ejected after one end of the specimen has been twisted through about 25° relative to the other, giving rise to a further 25° of plastic twist (Fig. 3).

We have assumed that the dislocation can glide in any plane containing its Burgers vector. In practice, certain crystallographic planes will be preferred. Such a plane will be itself twisted so that the way in which the dislocation leaves the rod will be rather complicated. Moreover, the movement of the original dislocation may trigger off more extensive plastic deformation. Still, we might expect that nonelastic behavior will begin at an angle of twist of about 25° (or, in general, $bl/2\pi R^2$).

CONCLUSIONS

Contrary to what one might expect, the image force on a screw dislocation parallel to the axis of a long rod tends to drive it towards the axis so long as its distance from the axis is less than 0.54 of the radius of the rod. Under deformations of the rod below a certain limit it will therefore not tend to leave the rod. Hence, from the increased strength of metallic whiskers we cannot immediately conclude that they are free from dislocations. Suitable torsion and flexion experiments might reveal their presence.

The author would like to thank Professor J. S. Koehler for helpful discussions.

[7] J. K. Mackenzie, thesis, Bristol (1949).

[8] Wilson, Research **2**, 541 (1949).

JOURNAL OF APPLIED PHYSICS VOLUME 24, NUMBER 2 FEBRUARY, 1953

Solution of Nonlinear Differential Equations by the Reversion Method

GEORGE I. COHN,* *Illinois Institute of Technology, Chicago, Illinois*

AND

BERNARD SALTZBERG,† *Engineering Laboratory, American Television Manufacturing Corporation, Chicago, Illinois*

(Received August 22, 1952)

By the use of the reversion method certain types of ordinary nonlinear differential equations can be reduced to a set of ordinary linear differential equations which may conveniently be solved by the Laplace transform. The expression for the general member of the linear set is developed and written out in detail for certain illustrative terms. The method is applied to two simple problems for which the answers are exactly known so that some idea may be obtained as to the excellence of the approximations supplied by this method.

I. INTRODUCTION

INTEREST in nonlinear equations has been expanding very rapidly due to the growing importance of phenomena which may not be represented by linear equations with sufficient accuracy. The reversion method which enables some ordinary nonlinear differential equations to be reduced to a set of ordinary linear differential equations is an important advancement in this field.[1,2] The purpose of this paper is (1) to reorganize the presentation of this method, (2) to develop a closed compact expression for the general differential equation of the reduced linear set, (3) to illustrate the method and the degree of approximation obtained by solving two simple problems whose solutions may also be obtained exactly so that some idea as to the rate of convergence may be obtained, and (4) to illustrate how this method may be extended to cover a broader variety of nonlinear differential equations.

II. GENERATION OF THE ASSOCIATED SET OF LINEAR DIFFERENTIAL EQUATIONS

The reversion method may be conveniently applied to equations of the type

$$Z_1I+Z_2I^2+Z_3I^3+\cdots+Z_nI^n=\lambda V(t), \tag{1}$$

where $I(t)$ is the unknown response to the given driving function $V(t)$ and the Z_k are functional forms of the differential and integral operators. The factor λ is an ordering parameter which facilitates the development of the associated linear set and is set equal to unity once it has served its purpose.

To reduce the nonlinear equation to a set of linear equations it is convenient to consider a solution of the form

$$I=I_1+I_2+\cdots+I_m+\cdots, \tag{2}$$

where each I_m depends only on the previous terms and I_1 is the solution of the original equation with all nonlinear terms absent. In order for the first term to be the response to the linear portion of the system and higher terms the additional response due to the nonlinearity the ordering parameter λ is introduced into the solution as follows:

$$I=\sum_{m=1}^{\infty}\lambda^mI_m. \tag{3}$$

Substituting (3) into (1) gives

$$\sum_{k=1}^{n}Z_k(\sum_{m=1}^{\infty}\lambda^mI_m)^k=\lambda V(t). \tag{4}$$

This equation is valid for any λ only if coefficients of like powers of λ on both sides of the equation are equal. Equating coefficients of like powers gives

$$Z_1I_1=V, \tag{5}$$

$$Z_1I_2=-Z_2I_1^2, \tag{6}$$

$$Z_1I_3=-Z_3I_1^3-2Z_2I_1I_2, \tag{7}$$

$$Z_1I_4=-Z_4I_1^4-3Z_3I_1^2I_2-Z_2(2I_1I_3+I_2^2), \tag{8}$$

$$Z_1I_5=-Z_5I_1^5-4Z_4I_1^3I_2-Z_3(3I_1^2I_3+3I_1I_2^2)-Z_2(2I_1I_4+2I_2I_3), \tag{9}$$

etc.,

which constitutes the associated set of linear differential equations. These equations are all identical except for the driving function. The driving function of each equation is given in terms of the solutions of the previous equations except for the first equation in which the driving function is that of the original nonlinear differential equation.

III. DERIVATION OF THE GENERAL DRIVING FUNCTION

The first few equations of the auxiliary set can easily be written down by inspection. However, each succeeding equation becomes lengthier and more involved. Therefore, a general expression for the driving functions would be useful. This may be obtained by putting all the nonlinear terms in Eq. (4) on the right-hand side, giving

$$Z_1\sum_{m=1}^{\infty}\lambda^mI_m=\lambda V(t)-\sum_{k=2}^{n}Z_k(\sum_{r=1}^{\infty}\lambda^rI_r)^k. \tag{10}$$

* Associate Professor of Electrical Engineering, Illinois Institute of Technology.

† Mathematical Consultant, Engineering Laboratory, American Television Manufacturing Corporation.

[1] S. Yèe, unpublished M.S. thesis, Illinois Institute of Technology (June, 1951).

[2] L. A. Pipes, J. Appl. Phys. 23, 202 (1952).

Equating corresponding powers of λ gives for the general equation of the associated linear set

$$Z_1 I_1 = V(t) \quad \text{for} \quad m=1, \tag{11}$$

$$Z_1 I_m = -\frac{1}{m!}\left[\frac{d^m}{d\lambda^m}\sum_{k=2}^{n} Z_k\Big(\sum_{r=1}^{\infty}\lambda^r I_r\Big)^k\right]_{\lambda=0} \quad m\geq 2. \tag{12}$$

This may be simplified to

$$Z_1 I_m = -\frac{1}{m!}\left[\frac{d^m}{d\lambda^m}\sum_{k=2}^{m} Z_k\Big(\sum_{r=1}^{m/2}\lambda^r I_r\Big)^k\right]_{\lambda=0} \quad m\geq 2, \tag{13}$$

since all higher terms in the summations contribute powers of λ which are higher than the mth and hence vanish. Re-expressing the inner summation to the kth power as follows:

$$\Big(\sum_{r=1}^{m}\lambda^r I_r\Big)^k = \sum_{j=k}^{mk/2} C_{jk}\lambda^j, \tag{14}$$

gives upon substitution into (13)

$$Z_1 I_m = -\frac{1}{m!}\left[\frac{d^m}{d\lambda^m}\sum_{k=2}^{m} Z_k \sum_{j=k}^{mk/2} C_{jk}\lambda^j\right]_{\lambda=0} \quad m\geq 2. \tag{15}$$

All powers of λ different from the mth vanish; hence

$$Z_1 I_m = -\frac{1}{m!}\left[\frac{d^m}{d\lambda^m}\sum_{k=2}^{m} Z_k C_{mk}\lambda^m\right]_{\lambda=0} \quad m\geq 2, \tag{16}$$

which immediately reduces to

$$Z_1 I_m = -\sum_{k=2}^{m} Z_k C_{mk} \quad m\geq 2. \tag{17}$$

C_{mk} is most easily determined by applying some number theoretical concepts to Eq. (14). Since by definition C_{mk} is the coefficient of λ^m in the right member of (14), we may utilize the identity between the power r of the ordering parameter and the subscript r of I_r in the left member of (14) to greatly facilitate the determination of C_{mk}. Note that the left member of (14) is the kth power of an mth degree polynomial in λ, and thus each coefficient in its expanded form will be the product of k coefficients of the original mth degree polynomial. It also follows from the correspondence of subscripts of the coefficients to the powers of λ in the mth degree polynomial that the coefficients of the expanded polynomial will be composed of k factors whose subscripts will sum to the power of λ it multiplies. Further, a particular term will be obtained as many times as the number of different ways the subscripts can be arranged. Thus, the problem of determining C_{mk} is replaced by the number theoretical one of generating all possible k-tuples of positive integers such that their sum is m, with each of the k-tuples taken as many times as the possible ways it can be arranged. For example, the number 15 can be written as the sum of 3 integers in the following ways:

(13, 1, 1) (12, 2, 1) (11, 3, 1) (11, 2, 2) (10, 4, 1)
(10, 3, 2) (9, 5, 1) (9, 4, 2) (9, 3, 3) (8, 6, 1)
(8, 5, 2) (8, 4, 3) (7, 7, 1) (7, 6, 2) (7, 5, 3)
(7, 4, 4) (6, 5, 4) (6, 6, 3) (5, 5, 5). (18)

The possible ways of arranging each of the above is given, respectively, by

$$\frac{3!}{2!},\ 3!,\ 3!,\ \frac{3!}{2!},\ 3!,\ 3!,\ 3!,\ 3!,\ \frac{3!}{2!},\ 3!,\ 3!,\ 3!,$$

$$\frac{3!}{2!},\ 3!,\ 3!,\ \frac{3!}{2!},\ 3!,\ \frac{3!}{2!},\ \frac{3!}{3!}. \tag{19}$$

The possible ways of arranging each of the k-tuples is given by the well-known formula

$$S_k!/S_1!S_2!\cdots S_n!, \tag{20}$$

where S is the number of objects in the set to be arranged and the S_j is the number of like elements of the jth kind. If all the elements are different the possible arrangements become $S!$.

From (18) and (19) it is now a simple matter to write

$$\begin{aligned} C_{15,3} = 3!\Big(&\frac{1}{2!}I_{13}I_1^2 + I_{12}I_2I_1 + I_{11}I_3I_1 + \frac{1}{2!}I_{11}I_2^2 \\ &+ I_{10}I_4I_1 + I_{10}I_3I_2 + I_9I_5I_1 + I_9I_4I_2 \\ &+ \frac{1}{2!}I_9I_3^2 + I_8I_6I_1 + I_8I_5I_2 + I_8I_4I_3 + \frac{1}{2!}I_7^2I_1 \\ &+ I_7I_6I_2 + I_7I_5I_3 + \frac{1}{2!}I_7I_4^2 + I_6I_5I_4 \\ &+ \frac{1}{2!}I_6^2I_3 + \frac{1}{3!}I_5^3\Big). \end{aligned} \tag{21}$$

For m large the total number of terms in the driving function of Eq. (17) is great, and it becomes convenient to have a check on the errors of omission. This check can be made with the following formula for the total number of terms appearing in each driving function:

Total number of terms

$$= \sum_{k=2}^{m} C_{k-1}^{m-1} = \sum_{k=2}^{m}\frac{(m-1)!}{(k-1)!(m-k)!}, \tag{22}$$

where $C_{k-1}{}^{m-1}$ is the number of combinations of $m-1$ different things taken $k-1$ at a time. The sum of the integers in (19) is a special case of this formula with $m=15$, and the contribution of $k=3$ which gives

$$C_{k-1}^{m-1} = C_2{}^{14} = \frac{14!}{2!12!} = 91. \tag{23}$$

The general formula (17) for the linear set will now be applied to extend the number of tabulated equations from thirteen[2] to fifteen. The 14th equation is given by

$$\begin{aligned} Z_1 I_{14} = &-Z_2 C_{14,2} - Z_3 C_{14,3} - Z_4 C_{14,4} - Z_5 C_{14,5} \\ &- Z_6 C_{14,6} - Z_7 C_{14,7} - Z_8 C_{14,8} - Z_9 C_{14,9} \\ &- Z_{10} C_{14,10} - Z_{11} C_{14,11} - Z_{12} C_{14,12} \\ &- Z_{13} C_{14,13} - Z_{14} C_{14,14}, \end{aligned} \tag{24}$$

where

$$C_{14,2} = 2!\left(I_{13}I_1 + I_{12}I_2 + I_{11}I_3 + I_{10}I_4 + I_9I_5 + I_8I_6 + \frac{1}{2!}I_7^2 \right), \tag{25}$$

$$C_{14,3} = 3!\left(\frac{1}{2!}I_{12}I_1^2 + I_{11}I_2I_1 + I_{10}I_3I_1 + \frac{1}{2!}I_{10}I_2^2 + I_9I_4I_1 + I_9I_3I_2 + I_8I_5I_1 + I_8I_4I_2 + \frac{1}{2!}I_8I_3^2 + I_7I_6I_1 + I_7I_5I_2 + I_7I_4I_3 + \frac{1}{2!}I_6^2I_2 + I_6I_5I_3 + \frac{1}{2!}I_6I_4^2 + \frac{1}{2!}I_5^2I_4 \right), \tag{26}$$

$$C_{14,4} = 4!\left(\frac{1}{3!}I_{11}I_1^3 + \frac{1}{2!}I_{10}I_2I_1^2 + \frac{1}{2!}I_9I_3I_1^2 + \frac{1}{2!}I_9I_2^2I_1 + I_8I_3I_2I_1 + \frac{1}{2!}I_8I_4I_1^2 + \frac{1}{3!}I_8I_2^3 + \frac{1}{2!}I_7I_5I_1^2 + I_7I_4I_2I_1 + \frac{1}{2!}I_7I_3^2I_1 + \frac{1}{2!}I_7I_3I_2^2 + \frac{1}{2!2!}I_6^2I_1^2 + I_6I_5I_2I_1 + I_6I_4I_3I_1 + \frac{1}{2!}I_6I_3^2I_2 + \frac{1}{2!2!}I_5^2I_2^2 + \frac{1}{2!}I_5^2I_3I_1 + \frac{1}{2!}I_5I_4^2I_1 + I_5I_4I_3I_2 + \frac{1}{3!}I_5I_3^3 + \frac{1}{3!}I_4^3I_2 + \frac{1}{2!2!}I_4^2I_3^2 \right), \tag{27}$$

$$C_{14,5} = 5!\left(\frac{1}{4!}I_{10}I_1^4 + \frac{1}{3!}I_9I_2I_1^3 + \frac{1}{3!}I_8I_3I_1^3 + \frac{1}{2!2!}I_8I_2^2I_1^2 + \frac{1}{3!}I_7I_4I_1^3 + \frac{1}{2!}I_7I_3I_2I_1^2 + \frac{1}{3!}I_7I_2^3I_1 + \frac{1}{3!}I_6I_5I_1^3 + \frac{1}{2!}I_6I_4I_2I_1^2 + \frac{1}{2!}I_6I_3^2I_1^2 + \frac{1}{2!}I_6I_3I_2^2I_1 + \frac{1}{4!}I_6I_2^4 + \frac{1}{2!}I_5I_4I_3I_1^2 + \frac{1}{2!}I_5I_4I_2^2I_1 + \frac{1}{2!}I_5I_3^2I_2I_1 + \frac{1}{3!}I_5I_3I_2^3 + \frac{1}{2!2!}I_5^2I_2I_1^2 + \frac{1}{2!3!}I_4^3I_1^2 + \frac{1}{3!}I_4I_3^2I_1 + \frac{1}{2!3!}I_4^2I_2^3 + \frac{1}{2!2!}I_4I_3^2I_2^2 + \frac{1}{2!}I_4^2I_3I_2I_1 + \frac{1}{4!}I_3^4I_2 \right), \tag{28}$$

$$C_{14,6} = 6!\left(\frac{1}{5!}I_9I_1^5 + \frac{1}{4!}I_8I_2I_1^4 + \frac{1}{4!}I_7I_3I_1^4 + \frac{1}{2!3!}I_7I_2^2I_1^3 + \frac{1}{4!}I_6I_4I_1^4 + \frac{1}{3!}I_6I_3I_2I_1^3 + \frac{1}{2!3!}I_6I_2^3I_1^2 + \frac{1}{2!4!}I_5^2I_1^4 + \frac{1}{3!}I_5I_4I_2I_1^3 + \frac{1}{2!2!}I_5I_3I_2^2I_1^2 + \frac{1}{2!3!}I_5I_3^2I_1^3 + \frac{1}{2!3!}I_4^2I_3I_1^3 + \frac{1}{2!2!2!}I_4^2I_2^2I_1^2 + \frac{1}{2!2!}I_4I_3^2I_2I_1^2 + \frac{1}{3!}I_4I_3I_2^3I_1 + \frac{1}{2!3!}I_3^3I_2^2I_1 + \frac{1}{2!4!}I_3^4I_1^2 + \frac{1}{5!}I_4I_2^5 + \frac{1}{2!4!}I_3^2I_2^4 \right), \tag{29}$$

$$C_{14,7} = 7!\left(\frac{1}{6!}I_8I_1^6 + \frac{1}{5!}I_7I_2I_1^5 + \frac{1}{5!}I_6I_3I_1^5 + \frac{1}{2!4!}I_6I_2^2I_1^4 + \frac{1}{5!}I_5I_4I_1^5 + \frac{1}{4!}I_5I_3I_2I_1^4 + \frac{1}{3!3!}I_5I_2^3I_1^3 + \frac{1}{2!4!}I_4I_3^2I_1^4 + \frac{1}{2!4!}I_4^2I_2I_1^4 + \frac{1}{3!3!}I_3^3I_2I_1^3 + \frac{1}{2!3!2!}I_3^2I_2^3I_1^2 + \frac{1}{5!}I_3I_2^5I_1 + \frac{1}{7!}I_2^7 + \frac{1}{2!4!}I_4I_2^4I_1^2 + \frac{1}{2!3!}I_4I_3I_2^2I_1^3 \right), \tag{30}$$

$$C_{14,8}=8!\Big(\frac{1}{7!}I_7I_1^7+\frac{1}{6!}I_6I_2I_1^6+\frac{1}{6!}I_5I_3I_1^6+\frac{1}{2!5!}I_5I_2^2I_1^5+\frac{1}{2!6!}I_4^2I_1^6+\frac{1}{5!}I_4I_3I_2I_1^5+\frac{1}{3!4!}I_4I_2^3I_1^4+\frac{1}{3!5!}I_3^3I_1^5+\frac{1}{2!2!4!}I_3^2I_2^2I_1^4+\frac{1}{4!3!}I_3I_2^4I_1^3+\frac{1}{6!2!}I_2^6I_1^2\Big), \tag{31}$$

$$C_{14,9}=9!\Big(\frac{1}{8!}I_6I_1^8+\frac{1}{7!}I_5I_2I_1^7+\frac{1}{7!}I_4I_3I_1^7+\frac{1}{2!6!}I_4I_2^2I_1^6+\frac{1}{2!6!}I_3^2I_2I_1^6+\frac{1}{3!5!}I_3I_2^3I_1^5+\frac{1}{5!4!}I_2^5I_1^4\Big), \tag{32}$$

$$C_{14,10}=10!\Big(\frac{1}{9!}I_5I_1^9+\frac{1}{8!}I_4I_2I_1^8+\frac{1}{2!7!}I_3I_2^2I_1^7+\frac{1}{2!8!}I_3^2I_1^8+\frac{1}{4!6!}I_2^4I_1^6\Big), \tag{33}$$

$$C_{14,11}=11!\Big(\frac{1}{10!}I_4I_1^{10}+\frac{1}{9!}I_3I_2I_1^9+\frac{1}{3!8!}I_2^3I_1^8\Big), \tag{34}$$

$$C_{14,12}=12!\Big(\frac{1}{11!}I_3I_1^{11}+\frac{1}{2!10!}I_2^2I_1^{10}\Big), \tag{35}$$

$$C_{14,13}=\frac{13!}{12!}I_2I_1^{12}, \tag{36}$$

$$C_{14,14}=\frac{14!}{14!}I_1^{14}. \tag{37}$$

The 15th equation is given by

$$\begin{aligned}Z_1I_{15}=&-Z_2C_{15,2}-Z_3C_{15,3}-Z_4C_{15,4}-Z_5C_{15,5}\\&-Z_6C_{15,6}-Z_7C_{15,7}-Z_8C_{15,8}-Z_9C_{15,9}\\&-Z_{10}C_{15,10}-Z_{11}C_{15,11}-Z_{12}C_{15,12}\\&-Z_{13}C_{15,13}-Z_{14}C_{15,14}-Z_{15}C_{15,15},\end{aligned} \tag{38}$$

where

$$C_{15,2}=2!(I_{14}I_1+I_{13}I_2+I_{12}I_3+I_{11}I_4+I_{10}I_5+I_9I_6+I_8I_7), \tag{39}$$

$$C_{15,3}=3!\Big(\frac{1}{2!}I_{13}I_1^2+I_{12}I_2I_1+\frac{1}{2!}I_{11}I_2^2+I_{11}I_3I_1+I_{10}I_4I_1+I_{10}I_3I_2+I_9I_5I_1+I_9I_4I_2+\frac{1}{2!}I_9I_3^2+I_8I_6I_1+I_8I_5I_2+I_8I_4I_3+\frac{1}{2!}I_7^2I_1+I_7I_6I_2+I_7I_5I_3+\frac{1}{2!}I_7I_4^2+I_6I_5I_4+\frac{1}{2!}I_6^2I_3+\frac{1}{3!}I_5^3\Big), \tag{40}$$

$$C_{15,4}=4!\Big(\frac{1}{3!}I_{12}I_1^3+\frac{1}{2!}I_{11}I_2I_1^2+\frac{1}{2!}I_{10}I_3I_1^2+\frac{1}{2!}I_{10}I_2^2I_1+\frac{1}{2!}I_9I_4I_1^2+I_9I_3I_2I_1+\frac{1}{3!}I_9I_2^3+\frac{1}{2!}I_8I_5I_1^2+I_8I_4I_2I_1+\frac{1}{2!}I_8I_3^2I_1+\frac{1}{2!}I_8I_3I_2^2+\frac{1}{2!}I_7I_6I_1^2+I_7I_4I_3I_1+\frac{1}{2!}I_7I_4I_2^2+\frac{1}{2!}I_7I_3^2I_2+I_7I_5I_2I_1+\frac{1}{2!}I_6I_5I_2^2+I_6I_4I_3I_2+\frac{1}{3!}I_6I_3^3+\frac{1}{2!}I_6^2I_2I_1+\frac{1}{2!}I_6I_4^2I_1+\frac{1}{2!}I_5^2I_4I_1+\frac{1}{2!}I_5^2I_3I_2+\frac{1}{2!}I_5I_4^2I_2+\frac{1}{2!}I_5I_3^2I_4+\frac{1}{3!}I_4^3I_3\Big), \tag{41}$$

$$C_{15,5}=5!\Big(\frac{1}{4!}I_{11}I_1^4+\frac{1}{3!}I_{10}I_2I_1^3+\frac{1}{3!}I_9I_3I_1^3+\frac{1}{2!2!}I_9I_2^2I_1^2+\frac{1}{3!}I_8I_4I_1^3+\frac{1}{2!}I_8I_3I_2I_1^2+\frac{1}{3!}I_8I_2^3I_1+\frac{1}{2!}I_7I_3I_2^2I_1+\frac{1}{3!}I_7I_5I_1^3+\frac{1}{2!}I_7I_4I_2I_1^2+\frac{1}{2!2!}I_7I_3^2I_1^2+\frac{1}{4!}I_7I_2^4+\frac{1}{2!}I_6I_3^2I_2I_1+\frac{1}{2!3!}I_6^2I_1^3+\frac{1}{2!}I_6I_5I_2I_1^2$$

$$+\frac{1}{2!}I_6I_4I_3I_1^2+\frac{1}{3!}I_6I_2^3I_3+\frac{1}{2!}I_6I_4I_2^2I_1$$
$$+\frac{1}{2!2!}I_5^2I_3I_1^2+\frac{1}{2!2!}I_5^2I_2^2I_1+\frac{1}{2!2!}I_5I_4^2I_1^2$$
$$+I_5I_4I_3I_2I_1+\frac{1}{3!}I_5I_4I_2^3+\frac{1}{3!}I_5I_3^2I_1$$
$$+\frac{1}{2!2!}I_5I_3^2I_2^2+\frac{1}{3!}I_4^3I_2I_1+\frac{1}{2!2!}I_4^2I_3^2I_1$$
$$+\frac{1}{2!2!}I_4^2I_3I_2^2+\frac{1}{2!}I_4I_3^3I_2+\frac{1}{5!}I_3^5\Big), \quad (42)$$

$$C_{15,6}=6!\Big(\frac{1}{5!}I_{10}I_1^5+\frac{1}{4!}I_9I_2I_1^4+\frac{1}{4!}I_8I_3I_1^4$$
$$+\frac{1}{2!3!}I_8I_2^2I_1^3+\frac{1}{4!}I_7I_4I_1^4+\frac{1}{3!}I_7I_3I_2I_1^3$$
$$+\frac{1}{2!3!}I_7I_2^3I_1^2+\frac{1}{4!}I_6I_5I_1^4+\frac{1}{2!3!}I_6I_3^2I_1^3$$
$$+\frac{1}{3!}I_6I_4I_2I_1^3+\frac{1}{2!2!}I_6I_3I_2^2I_1^2+\frac{1}{4!}I_6I_2^4I_1$$
$$+\frac{1}{2!3!}I_5^2I_2I_1^3+\frac{1}{2!2!}I_5I_3^2I_2I_1^2+\frac{1}{3!}I_5I_4I_3I_1^3$$
$$+\frac{1}{3!}I_5I_3I_2^3I_1+\frac{1}{5!}I_5I_2^5+\frac{1}{2!2!}I_5I_4I_2^2I_1^2$$
$$+\frac{1}{3!3!}I_4^3I_1^3+\frac{1}{2!3!}I_4^2I_2^3I_1+\frac{1}{2!3!}I_4I_3^3I_1^2$$
$$+\frac{1}{2!2!}I_4^2I_3I_2I_1^2+\frac{1}{2!2!}I_4I_3^2I_2^2I_1+\frac{1}{4!}I_4I_3I_2^4$$
$$+\frac{1}{4!}I_3^4I_2I_1+\frac{1}{3!3!}I_3^3I_2^3\Big), \quad (43)$$

$$C_{15,7}=7!\Big(\frac{1}{6!}I_9I_1^6+\frac{1}{5!}I_8I_2I_1^5+\frac{1}{2!4!}I_7I_2^2I_1^4$$
$$+\frac{1}{5!}I_7I_3I_1^5+\frac{1}{5!}I_6I_4I_1^5+\frac{1}{4!}I_6I_3I_2I_1^4$$
$$+\frac{1}{3!3!}I_6I_2^3I_1^3+\frac{1}{2!5!}I_5^2I_1^5+\frac{1}{2!4!}I_5I_3^2I_1^4$$
$$+\frac{1}{4!}I_5I_4I_2I_1^4+\frac{1}{2!3!}I_5I_3I_2^2I_1^3+\frac{1}{2!4!}I_5I_2^4I_1^2$$
$$+\frac{1}{2!4!}I_4^2I_3I_1^4+\frac{1}{2!2!3!}I_4^2I_2^2I_1^3+\frac{1}{2!3!}I_4I_3^2I_2I_1^3$$
$$+\frac{1}{5!}I_4I_2^5I_1+\frac{1}{2!3!}I_4I_3I_2^3I_1^2+\frac{1}{6!}I_3I_2^6$$
$$+\frac{1}{2!4!}I_3^2I_2^4I_1+\frac{1}{2!2!3!}I_3^3I_2^2I_1^2+\frac{1}{3!4!}I_3^4I_1^3\Big), \quad (44)$$

$$C_{15,8}=8!\Big(\frac{1}{7!}I_8I_1^7+\frac{1}{6!}I_7I_2I_1^6+\frac{1}{6!}I_6I_3I_1^6$$
$$+\frac{1}{2!5!}I_6I_2^2I_1^5+\frac{1}{6!}I_5I_4I_1^6+\frac{1}{5!}I_5I_3I_2I_1^5$$
$$+\frac{1}{3!4!}I_5I_2^3I_1^4+\frac{1}{2!5!}I_4^2I_2I_1^5+\frac{1}{2!5!}I_4I_3^2I_1^5$$
$$+\frac{1}{3!4!}I_4I_2^4I_1^3+\frac{1}{2!4!}I_4I_3I_2^2I_1^4+\frac{1}{3!4!}I_3^3I_2I_1^4$$
$$+\frac{1}{2!5!}I_3I_2^5I_1^2+\frac{1}{2!3!3!}I_3^2I_2^3I_1^3+\frac{1}{7!}I_2^7I_1\Big), \quad (45)$$

$$C_{15,9}=9!\Big(\frac{1}{8!}I_7I_1^8+\frac{1}{7!}I_6I_2I_1^7+\frac{1}{7!}I_5I_3I_1^7$$
$$+\frac{1}{2!6!}I_5I_2^2I_1^6+\frac{1}{2!7!}I_4^2I_1^7+\frac{1}{6!}I_4I_3I_2I_1^6$$
$$+\frac{1}{3!5!}I_4I_2^3I_1^5+\frac{1}{3!6!}I_3^2I_1^6+\frac{1}{4!4!}I_3I_2^4I_1^4$$
$$+\frac{1}{2!2!5!}I_3^2I_2^2I_1^5+\frac{1}{3!6!}I_2^6I_1^3\Big), \quad (46)$$

$$C_{15,10}=10!\Big(\frac{1}{9!}I_6I_1^9+\frac{1}{8!}I_5I_2I_1^8+\frac{1}{8!}I_4I_3I_1^8$$
$$+\frac{1}{2!7!}I_4I_2^2I_1^7+\frac{1}{2!7!}I_3^2I_2I_1^7$$
$$+\frac{1}{3!6!}I_3I_2^3I_1^6+\frac{1}{5!5!}I_2^5I_1^5\Big), \quad (47)$$

$$C_{15,11}=11!\Big(\frac{1}{10!}I_5I_1^{10}+\frac{1}{9!}I_4I_2I_1^9+\frac{1}{2!9!}I_3^2I_1^9$$
$$+\frac{1}{2!8!}I_3I_2^2I_1^8+\frac{1}{4!7!}I_2^4I_1^7\Big), \quad (48)$$

$$C_{15,12}=12!\Big(\frac{1}{11!}I_4I_1^{11}+\frac{1}{10!}I_3I_2I_1^{10}+\frac{1}{3!9!}I_2^3I_1^9\Big), \quad (49)$$

$$C_{15,13}=13!\left(\frac{1}{12!}I_3I_1^{12}+\frac{1}{2!11!}I_2^2I_1^{11}\right), \tag{50}$$

$$C_{15,14}=\frac{14!}{13!}I_2I_1^{13}, \tag{51}$$

$$C_{15,15}=\frac{15!}{15!}I_1^{15}. \tag{52}$$

IV. ILLUSTRATIVE EXAMPLES

Consider an initially quiescent circuit composed of a nonlinear element in series with an inductance, battery, and a switch as shown in Fig. 1. As a first example consider a nonlinear element with a characteristic such that the voltage drop across it is proportional to the square of the current. The equation characterizing the system is then

$$L\frac{dI}{dt}+RI^2=V. \tag{53}$$

To facilitate graphing the solution and its successive approximations obtained from the reversion method the variables are normalized. The normalized current is

$$i=\left(\frac{R}{V}\right)^{\frac{1}{2}}I, \tag{54}$$

and the normalized time is

$$\tau=\frac{(RV)^{\frac{1}{2}}}{L}t. \tag{55}$$

In terms of the normalized variables the equation is

$$\frac{di}{d\tau}+i^2=1. \tag{56}$$

In this example

$$Z_1=d/d\tau, \tag{57}$$

$$Z_2=1, \tag{58}$$

$$Z_k=0 \quad \text{for} \quad k\geq 2. \tag{59}$$

Therefore, the associated set of linear differential equations are

$$di_1/d\tau=1, \tag{60}$$

$$di_2/d\tau=-i_1^2, \tag{61}$$

$$di_3/d\tau=-2i_1i_2, \tag{62}$$

$$di_4/d\tau=-2i_1i_3-i_2^2, \tag{63}$$

$$di_5/d\tau=-2i_1i_4-2i_2i_3, \tag{64}$$

etc.

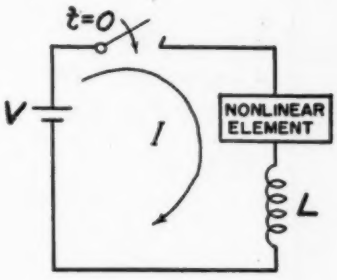


FIG. 1. Nonlinear circuit.

Solving these equations in succession gives

$$i_1=\tau, \tag{65}$$

$$i_2=-\tfrac{1}{3}\tau^3, \tag{66}$$

$$i_3=\frac{2}{15}\tau^5, \tag{67}$$

$$i_4=-\frac{17}{45}\tau^7, \tag{68}$$

$$i_m=(-1)^{m-1}\frac{2^{2m}(2^{2m}-1)}{(2m)!}B_{2m-1}\tau^{2m-1}, \tag{69}$$

where the B_i are Bernoulli numbers. The exact solution of (56) is

$$i=\tanh\tau. \tag{70}$$

Note that $i_1, i_2, i_3, \cdots$ are just the successive terms in the power series expansion of the exact solution when $|\tau|<(\pi/2)$. The successive approximations of the solution together with the exact solution are plotted for comparison in Fig. 2.

As a second example consider a nonlinear element such as a germanium crystal rectifier with a characteristic such that the current is proportional to the square of the voltage. The equation characterizing the system is then

$$L\frac{dI}{dt}+RI^{\frac{1}{2}}=V. \tag{71}$$

The normalized current in this example is

$$i=\frac{R^2}{V^2}I, \tag{72}$$

and the normalized time is

$$\tau=\frac{R^2}{LV}t. \tag{73}$$

In terms of the normalized variables the differential equation is

$$\frac{di}{d\tau}+i^{\frac{1}{2}}=1. \tag{74}$$

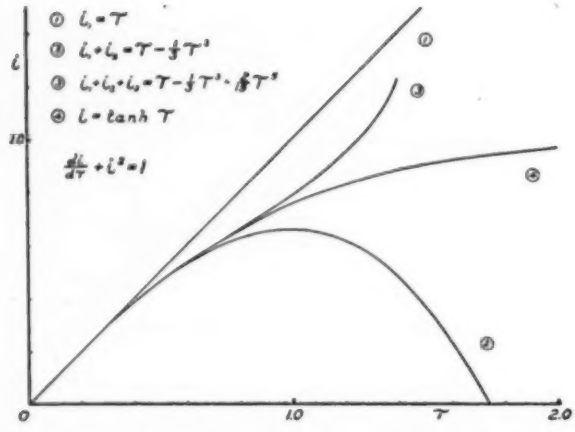


FIG. 2. Approximations and exact solution of $(di/d\tau)+i^2=1$.

This equation does not fit directly into the class of equations of the form given by (1) since a nonintegral power of the dependent variable occurs. At first it might be thought that this equation could be fitted into the scheme of Eq. (1) by rearranging (74) to remove the radical, thereby giving

$$2(di/d\tau)+i-(di/d\tau)^2=1. \tag{75}$$

However, the term $(di/d\tau)^2$ does not seem to be derivable from a form of the type Z_2i^2 because of the introduction of undesired terms which cannot be eliminated. Equation (74) may be brought into the scope of this paper by expanding the radical in a power series. The radical cannot be expanded about zero but may be conveniently expanded about unity. Rewriting the radical in a form suitable for such an expansion gives

$$(di/d\tau)+[1-(1-i)]^{\frac{1}{2}}=1. \tag{76}$$

Since the expansion will be in powers of $(1-i)$, it is more convenient to work with the variable

$$x=1-i, \tag{77}$$

in terms of which the differential Eq. (76) becomes

$$(dx/d\tau)-(1-x)^{\frac{1}{2}}=-1. \tag{78}$$

Expansion of the radical is

$$(1-x)^{\frac{1}{2}}=1-\sum_{k=1}^{\infty}\frac{(2k-3)!!}{2^k k!}x^k, \tag{79}$$

where

$$(2k-3)!!=(2k-3)(2k-5)!!, \tag{80}$$

with

$$(-1)!!=1. \tag{81}$$

Substitution of (79) into (78) gives

$$\frac{dx}{d\tau}+\sum_{k=1}^{\infty}\frac{(2k-3)!!}{2^k k!}x^k=0. \tag{82}$$

In this example

$$Z_1=\frac{d}{d\tau}+\tfrac{1}{2}, \tag{83}$$

$$Z_k=\frac{(2k-3)!!}{2^k k!}\quad k\geq 2. \tag{84}$$

Therefore the associated differential equations are

$$\frac{dx_1}{d\tau}+\frac{x_1}{2}=0, \tag{85}$$

$$\frac{dx_2}{d\tau}+\frac{x_2}{2}=-\frac{x_1^2}{8}, \tag{86}$$

etc.

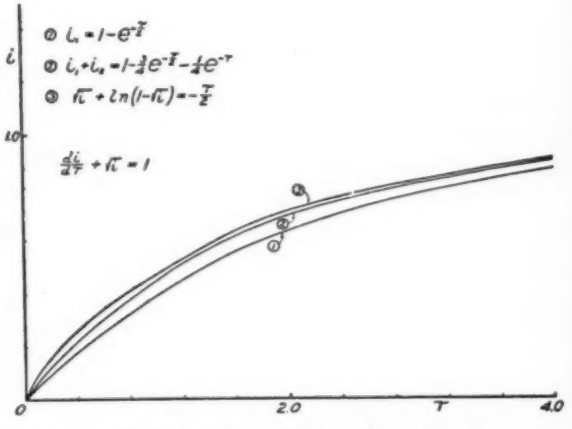


FIG. 3. Approximations and exact solution of $(di/d\tau)+\sqrt{i}=1$.

Solving these equations in succession gives

$$x_1=e^{-\tau/2}, \tag{87}$$

$$x_2=\tfrac{1}{4}(e^{-\tau}-e^{-\tau/2}), \tag{88}$$

etc.,

but

$$i=1-x \tag{89}$$

and hence

$$\sum_{k=1}^{n}i_k=1-\sum_{k=1}^{n}x_k. \tag{90}$$

The exact solution of this equation is

$$\ln(1-i^{\frac{1}{2}})+i^{\frac{1}{2}}=-\tau/2. \tag{91}$$

The exact solution together with the successive approximations for this solution are plotted in Fig. 3 for comparison.

JOURNAL OF APPLIED PHYSICS VOLUME 24, NUMBER 2 FEBRUARY, 1953

Quasi-Degenerated Modes in High-ϵ Dielectric Cavities

H. M. SCHLICKE*
Allen-Bradley Company, Milwaukee, Wisconsin
(Received September 26, 1952)

This paper deals with resonance effects in circularly cylindrical solid bodies of high-ϵ dielectrics such as titanates, these bodies being either not metallized or partly or fully metallized by fired-on silver. The high dielectric constant has two pronounced effects:

(a) Miniaturization, a linear reduction in size in the order of magnitude of 1/100, and
(b) Realization of boundary conditions approaching infinite wave impedance.

All four combinations of radial and axial wave impedance being zero (metallized) or infinite (interface dielectric/air) are investigated mathematically and experimentally. Special consideration is given to degenerated modes; that is, modes for which the axial dimensions of the cavity do not contribute to the resonance wavelength. Because of the inherent magnetic leakage through the dielectric, this is only conditionally true. These modes are therefore termed quasi-degenerated.

In contrast to conventional cavity theory dealing only with metallic boundaries, quasi-degenerated *TE* modes are realizable; namely, if the faces of the cylindrical cavities are not metallized. These cavities can be tuned by introducing magnetic rods into them or by application of conductive disks outside of them in the proper position.

A novel type of antenna can be evolved from the $TE_{010}^{\infty\infty}$ mode, comprising a dielectric spiral encircling a magnetic rod.

In many cases the relatively high temperature coefficient, dissipation factor, and noise level prevent for the time being the potential application of these high-ϵ dielectrics.

1. INTRODUCTION

In certain titanates or other materials having dielectric constants in the order of 2500 to 10 000, the wavelengths are only 1/50 to 1/100 of the wavelengths in air.

This highly miniaturizing effect transfers the microwave technique into the VHF range and in addition offers novel possibilities not obtainable with conventional "plumbing" that deals only with metallic boundaries constituting zero impedance. This boundary condition is reversed for nonmetallized high-ϵ dielectrics, since for a wave coming from inside the electrically "dense" material of high ϵ to the electrically "thin" air, the interface dielectric/air represents an infinite impedance. The higher the dielectric constant the more this idealization is approximated.

The starting Eqs. (1) to (6) in Sec. 2.1 are taken from Borgnis.[1] These relations follow from an extension of fundamental principles of propagation along definable cylinders as set forth by Lord Raleigh,[2] Honduras and Debye,[3] Schriever,[4] Carson, Mead, and Schelkunoff,[5] and others. Richtmeyer[6] investigated spherical and toroidal dielectric cavities. The latter case, though approached differently, is somewhat related to one of the three basically new types of degenerated modes that can exist in addition to the degenerated *TM* type realizable for metallic boundaries alone. Most of the work done previously on resonant cavities, however, has been restricted to configurations having metallic boundaries.[7-9]

The present paper, then, covers the field of solid cylindrical bodies of high-ϵ material in a systematic approach, thereby obtaining additional solutions and modifications not known so far.

2. THEORY

2.1 General Theory

The resonance wavelength λ' excitable in a cylindrical cavity of diameter D and length L is given as

$$\lambda' = 2\pi/K \tag{1}$$

in which

$$K^2 = K_1^2 + K_3^2, \tag{2}$$

where

$$K_1 = x_{lm}/(D/2) \quad \text{or} \quad x_{lm}'/(D/2) \tag{3}$$

with

$$x_{lm} = m\text{th root of } J_l(K_1 D/2) \tag{4}$$

or

$$x_{lm}' = m\text{th root of } J_l'(K_1 D/2) \tag{5}$$
$$l = 0, 1, 2, 3, \ldots$$

and where

$$K_3 = n\pi/L; \quad n = 0, 1, 2, 3, 4, \ldots. \tag{6}$$

$J_l(x)$ or $J_l'(x)$ are the Bessel functions of the first kind or their first derivatives, respectively. In case the cavity is filled with a medium of relative dielectric constant ϵ and relative permeability μ, the interrelation between frequency f and wavelength λ' (in the medium) is given by

$$\frac{3\times 10^{10} \text{ cm sec}^{-1}}{(\mu\cdot\epsilon)^{\frac{1}{2}}} = f\cdot\lambda'. \tag{7}$$

Using the impedance concept introduced by Schelkunoff, the metallic boundaries determine

$$\left.\begin{aligned} Z_z &= 0 \\ Z_r &= 0 \end{aligned}\right\} \text{at metallic boundary.} \tag{8a, 8b}$$

* Dr. H. M. Schlicke is Consultant Engineer with the Allen-Bradley Company, Milwaukee, Wisconsin.
[1] F. Borgnis, Ann. Physik **35**, 359 (1939).
[2] Lord Raleigh, Phil. Mag. **45**, 125–132 (1897).
[3] D. Honduras and P. Debye, Ann. Physik **32**, 465 (1910).
[4] W. Schriever, Ann. Physik **63**, 645 (1920).
[5] Carson, Mead, and Schelkunoff, Bell System Tech. J. **15**, 310 (1936).
[6] R. D. Richtmeyer, J. Appl. Phys. **10**, 391 (1939).
[7] W. L. Barrow and W. W. Mieher, Proc. Inst. Radio Engrs. **28**, 189 (1940).
[8] T. P. Kinzer, BTL, Case 23458, January 8, 1943.
[9] M.I.T. Rad. Lab. Series (McGraw-Hill Book Company, Inc., New York, 1947), Vol. 11, p. 297.

TABLE I. Survey of equations conditioned and correlated by modes and boundary conditions.

	a	b	c	d
11	$Z_z^+ = 0$		$Z_z^+ = \infty$	
12	$Z_z = \frac{K_3}{K} \tan K_3 z \; (= \frac{E_r}{H_\phi})$	$Z_z = \frac{K}{K_3} \tan K_3 z$	$Z_z = \frac{K_3}{K} \cot K_3 z$	$Z_z = \frac{K}{K_3} \cot K_3 z$
13	$= 0$ if $n = 0$	$\neq 0$ for $n = 0$	$\neq \infty$ for $n = 0$	$= \infty$ for $n = 0$
14		$\therefore n \neq 0$	$\therefore n \neq 0$	
15	TM_{-}^{0-} mode	TE_{-}^{0-} mode	$TM_{-}^{\infty-}$ mode	$TE_{-}^{\infty-}$ mode
16	$E_r = -\frac{K_3}{K} J'_\ell(K_1 r) \cos \ell\phi \sin K_3 z$	$E_r = -\ell \frac{J_\ell(K_1 r)}{K_1 r} \sin \ell\phi \sin K_3 z$	$E_r = -\frac{K_3}{K} J'_\ell(K_1 r) \cos \ell\phi \cos K_3 z$	$E_r = -\ell \frac{J_\ell(K_1 r)}{K_1 r} \sin \ell\phi \cos K_3 z$
17	$E_\phi = \ell \frac{K_3}{K} \frac{J_\ell(K_1 r)}{K_1 r} \sin \ell\phi \sin K_3 z$	$E_\phi = -J'_\ell(K_1 r) \cos \ell\phi \sin K_3 z$	$E_\phi = \ell \frac{K_3}{K} \frac{J_\ell(K_1 r)}{K_1 r} \sin \ell\phi \cos K_3 z$	$E_\phi = -J'_\ell(K_1 r) \cos \ell\phi \cos K_3 z$
18	$E_z = \frac{K_1}{K} J_\ell(K_1 r) \cos \ell\phi \cos K_3 z$	$E_z = 0$	$E_z = \frac{K_1}{K} J_\ell(K_1 r) \cos \ell\phi \sin K_3 z$	$E_z = 0$
19	$H_r = -\ell \frac{J_\ell(K_1 r)}{K_1 r} \sin \ell\phi \cos K_3 z$	$H_r = \frac{K_3}{K} J'_\ell(K_1 r) \cos \ell\phi \cos K_3 z$	$H_r = -\ell J'_\ell(K_1 r) \sin \ell\phi \sin K_3 z$	$H_r = \frac{K_3}{K} J'_\ell(K_1 r) \cos \ell\phi \sin K_3 z$
20	$H_\phi = -J'_\ell(K_1 r) \cos \ell\phi \cos K_3 z$	$H_\phi = \ell \frac{K_3}{K} \frac{J'_\ell(K_1 r)}{K_1 r} \sin \ell\phi \cos K_3 z$	$H_\phi = -J'_\ell(K_1 r) \cos \ell\phi \sin K_3 z$	$H_\phi = -\ell \frac{K_3}{K} \frac{J_\ell(K_1 r)}{K_1 r} \sin \ell\phi \sin K_3 z$
21	$H_z = 0$	$H_z = \frac{K_1}{K} J_\ell(K_1 r) \cos \ell\phi \sin K_3 z$	$H_z = 0$	$H_z = \frac{K_1}{K} J_\ell(K_1 r) \cos \ell\phi \cos K_3 z$
22	$Z_r = \frac{E_z}{H_\phi} = -\frac{K_1}{K} \frac{J_\ell(K_1 r)}{J'_\ell(K_1 r)}$	$Z_r = \frac{E_\phi}{H_z} = -\frac{K}{K_1} \frac{J'_\ell(K_1 r)}{J_\ell(K_1 r)}$	$Z_r = \frac{E_z}{H_\phi} = -\frac{K_1}{K} \frac{J_\ell(K_1 r)}{J'_\ell(K_1 r)}$	$Z_r = \frac{E_\phi}{H_z} = -\frac{K}{K_1} \frac{J'_\ell(K_1 r)}{J_\ell(K_1 r)}$
23	If $Z_r = 0 \longrightarrow J_\ell(K_1 r) = 0$	If $Z_r = 0 \longrightarrow J'_\ell(K_1 r) = 0$	If $Z_r = 0 \longrightarrow J_\ell(K_1 r) = 0$	If $Z_r = 0 \longrightarrow J'_\ell(K_1 r) = 0$
24	If $Z_r = \infty \longrightarrow J'_\ell(K_1 r) = 0$	If $Z_r = \infty \longrightarrow J_\ell(K_1 r) = 0$	If $Z_r = \infty \longrightarrow J'_\ell(K_1 r) = 0$	If $Z_r = \infty \longrightarrow J_\ell(K_1 r) = 0$
	Lowest degenerated mode for $Z_r = \infty$ (25) and $Z_r = 0$ (26)			
25	$TM_{110}^{0\,\infty\,\infty}$: $D/\lambda = \frac{1.83}{\pi}$ $L \to 0$	—	—	$TE_{010}^{\infty\,\infty\,\infty}$: $D/\lambda = \frac{2.4}{\pi}$ $L \to \infty$
26	$TM_{010}^{0\,0\,\infty}$: $D/\lambda = \frac{2.4}{\pi}$			$TE_{110}^{\infty\,\infty\,0}$: $D/\lambda = \frac{1.83}{\pi}$ $L \to \infty$

Since
fac
E_r
or
H_r
wh

or

in
equ
tio
The
tio
app
I
con
met

for
the
F
mat

or
whi

Hen
(16t
(19c
T
give
T
and
root
corr
(23d
dens
In
for

The
whe
(0)
nota

Singling out Z_z, the wave impedance of the cylinder face, there must be

$$E_t=0 \text{ (tangential comp. of el. field strength)} \quad (9a)$$

or

$$H_n=0 \text{ (normal component of magn. field strength)}, \quad (9b)$$

which requires

$$\left.\begin{aligned} E_r&=0 \\ E_\phi&=0 \end{aligned}\right\} \text{at the metallic boundary} \quad \begin{aligned}&(10a)\\&(10b)\end{aligned}$$

$$\text{or} \quad H_z=0 \quad (10c)$$

in cylindrical coordinates.

These field components must satisfy Maxwell's equations and are specified more completely by equations appearing in columns (a) and (b) of Table I. They may be contrasted with the corresponding equations for nonmetallized faces of the dielectric cylinder appearing in columns (c) and (d) of the same table.

If the cavity consists of a material of a high dielectric constant, and if the faces of the cylinder are not metallized,

$$Z_z \to \infty \quad (27)$$

for a wave looking from inside of the dielectric toward the air boundary.

For high ϵ this idealization is quite closely approximated. Eq. (27) conditions

$$H_t=0 \quad (28a)$$

or

$$E_n=0, \quad (28b)$$

which requires

$$\left.\begin{aligned} H_r&=0 \\ H_\phi&=0 \\ E_z&=0 \end{aligned}\right\} \text{at the boundary dielectric/air.} \quad \begin{aligned}&(29a)\\&(29b)\\&(29c)\end{aligned}$$

Hence the term $\sin K_z z$ previously seen in Eqs. (16a), (16b), (17a), and (17b) appears now in Eqs. (18c), (19c), (19d), (20c), and (20d). See Table I.

The impedance of the periphery of the cylinder Z_r is given by Eqs. (22a), (22b), (22c), and (22d).

Three parameters; namely, the type of mode, Z_z, and Z_r, are involved when it is to be decided whether roots of $J_l(K_1 r)=0$ or $J_l'(K_1 r)=0$ are to be used. The correlation is expressed in Eqs. (23a), (23b), (23c), (23d), and (24a), (24b), (24c), (24d) or, more condensed, in the table inserted in Fig. 1.

In the general case, a total of 5 indices are required for which the following notations were chosen:

$$TM_{lmn}^{Z_z Z_r} \quad \text{or} \quad TE_{lmn}^{Z_z Z_r}.$$

The superscripts will be either zero or ∞, depending on whether the face (Z_z) or the periphery (Z_r) is metallized (0) or not (∞). The subscripts follow the standard notation.[9]

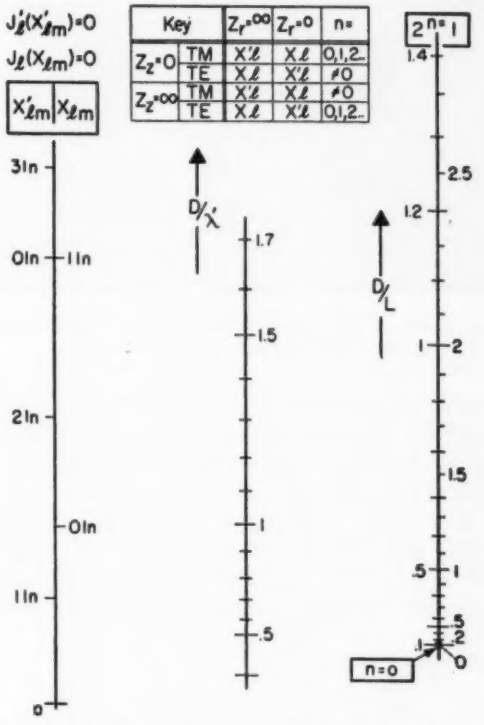


FIG. 1. Generalized mode nomograph for cylindrical cavities.

The generalized mode nomograph, Fig. 1, is based on

$$\left(\frac{D}{\lambda'}\right)^2=\left(\frac{x_{lm}^{(\prime)}}{\pi}\right)^2+\left(\frac{nD}{2L}\right)^2 \quad (30)$$

derived from Eqs. (1) to (6).

The right scale is a double one, for $n=1$, and $n=2$, where n denotes the number of halfwave variations with respect to z. $n=0$ is marked by a single point. The center scale signifies D/λ'. The left-hand line is the locus of a double set of specific points directly denoted by properly paired indices l and m. Virtually, these points give the zeros of the Bessel functions and of their first derivatives, both of the first kind. The table on the nomograph indicates which scale is to be read, depending on the mode and boundary conditions.

2.2 Quasi-Degenerated Modes

A mode is termed degenerated if the field component in the z direction is invariant with z, or equivalently, if $n=0$. If $Z_z=0$, only TM modes with $n=0$ can exist. This is well known. Compare Eqs. (12a) and (12b) and (13a) and (13b) in Table I.

However, in the case of $Z_z=\infty$, only TE modes with $n=0$ are realizable, as Eqs. (12d) and (13d) indicate. According to Eqs. (12c), (13c), and (14c) degenerated TM modes cannot be realized, if the face of the cylinder is not metallized.

There are four possible cases of degenerated mode configurations. The lowest of each group are: $TM_{110}^{0\infty}$, TM_{010}^{00}, $TE_{010}^{\infty\infty}$, and $TE_{110}^{\infty 0}$, where only the TM_{010}^{00}

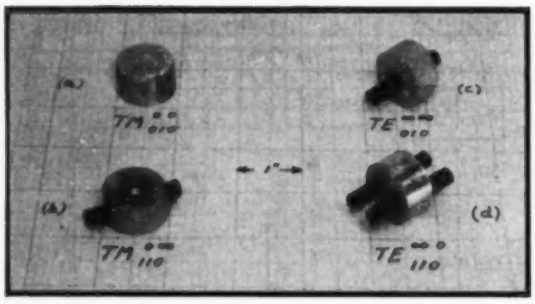


FIG. 2. The 4 basic types of lowest (quasi-) degenerated modes with magnetic tuning rods added.

is actually independent of L. In the three other mode types, strong magnetic leakage is inherent. All solutions of the underlying Bessel equations (Compare Bibliograph 1) are premised on the applicability of a cylindrical coordinate system. For TM modes ($Z_z=0$), this holds only if $L \ll D$; for TE modes ($Z_Z=\infty$), this holds only if $L \gg D$. Or, e.g., the theoretically calculated $TE_{010}^{\infty\infty}$ mode (lowest mode of a nonmetallized dielectric body of high ϵ) is only determinable by extrapolating to $L \gg D$, (so that leakage from the faces of the cylinder is negligible). Correspondingly, in the case of $TM_{110}^{0\infty}$ mode (high ϵ disk capacitor without leads), where the magnetic leakage occurs through two opposite sides of the circumference of the cylinder, the calculations are only exactly valid for L approaching 0. These modes could, therefore, be termed quasi-degenerated.

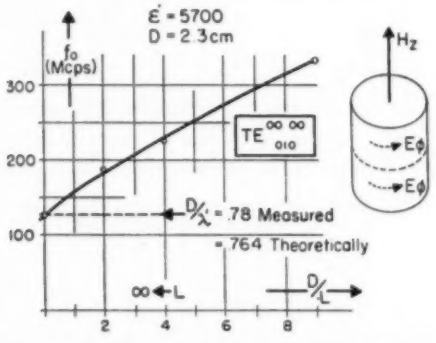


FIG. 3. Resonance frequency of nonmetallized dielectric cylinder as function of D/L for constant D.

These findings are of particular interest when applied to the calculation of the resonance frequency of circular disk condensers. King[10] finds that the lowest resonance frequency of a lead-less circular condenser is given by the equation

$$J_0(K_1 r)=0; \quad x_{01}=2.405 \tag{31}$$

denoting a TM_{010} mode. If, however, the dielectric of this lead-less capacitor has a high dielectric constant, the TM_{110} mode is excitable, conditioned by

$$J_1'(K_1 r)=0; \quad x_{11}'=1.841. \tag{32}$$

This means that the highest achievable resonance frequency is more than 30 percent lower for the capacitor with the high dielectric constant, as compared with the correspondingly larger capacitor having the same capacity but a low dielectric. It is understood that in both cases $D \gg L$.

Figure 2 shows all four basic types of lowest quasi-degenerated modes with magnetic tuning rods added where the magnetic field is strongest.

(a) In the upper left corner is the TM_{010}^{00} mode, well known from conventional cavity theory, except that the cavities are miniaturized because of the high dielectric constant. The scale of the paper square is one inch. The resonance frequency is in the order of 100 Mcps for an ϵ of 5700.

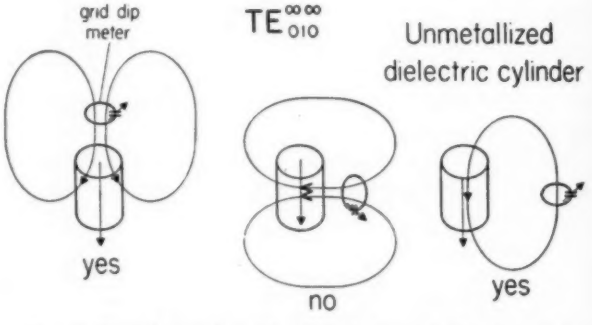


FIG. 5. Experimental verification of theoretically obtained field configurations of $TE_{010}^{\infty\infty}$ mode.

(b) In the lower left corner is the case of the high-ϵ dielectric condenser, the $TM_{110}^{0\infty}$ mode that was already discussed. A ferrite rod used for tuning indicates the basic arrangement of the magnetic field.

(c) In the upper right corner is the quasi-degenerated $TE_{010}^{\infty\infty}$ mode, a cylindrical dielectric cavity without any metallized surface. The resonance frequency is lowered by a magnetic rod, encircled by the electric field in the dielectric.

(d) The lower right corner depicts the $TE_{110}^{\infty 0}$ mode, the complementary one to the $TM_{110}^{0\infty}$ mode (b). Here the boundary conditions are reversed and consequently the electric and the magnetic fields are interchanged. This is illustrated by a pair of ferrite tuning rods. In these rods the magnetic displacement current flows in opposite directions, since both rods form part of a closed magnetic circuit.

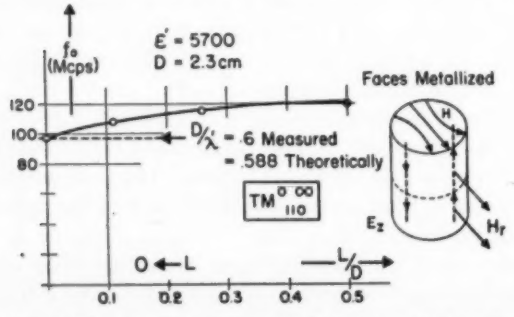


FIG. 4. Resonance frequency of dielectric cylinder with metallized faces as function of L/D for D=constant.

[10] R. W. P. King, *Electromagnetic Waves* (McGraw-Hill Book Company, Inc., New York, 1945), Vol. 1, p. 367.

The magnetic tuning rods used permitted a tuning range of about 25 percent for the $TE_{010}^{\infty\infty}$ and the $TM_{110}^{0\infty}$ modes, and more for the $TE_{110}^{\infty 0}$ mode. Tuning by means of ferrite rods cannot be applied above the vhf range because of unavoidable gyromagnetic resonance absorption. However, in the uhf range where dielectric constants between 100 and 1000 lead to conveniently small dimensions for cavities, the tuning can be achieved effectively by using copper or brass disks or rings in the proper position close to the cavity. Then the cavities need not contain any holes, since the external magnetic field is sufficiently strong for coupling. Slots and holes in the proper position interrupting the electric field may also be used for changing the resonance frequency or for suppressing higher modes. The magnetic leakage of the cavities provides magnetic coupling to the associated circuits without mechanical contact. This is important for possible applications requiring switching.

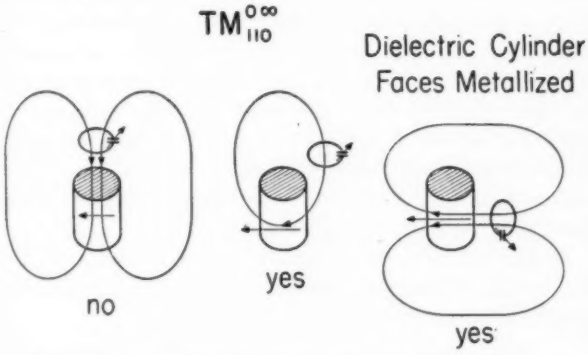


FIG. 6. Experimental verification of theoretically obtained field configuration of $TM_{110}^{0\infty}$ mode.

3. EXPERIMENTAL VERIFICATION

3.1 Measurements on the Lowest Quasi-Degenerated Modes

Experimental verification was made with the cylindrical bodies of 2.3 cm diameter and various lengths, the bodies having dielectric constants of 5700, 3300, and 2000, respectively. Fired-on silver was used for metallic boundaries. The measuring device applied was a grid-dip-meter (Megacycle-Meter made by Measurements Corporation, Boonton, New Jersey). All tests were based on magnetic coupling with the oscillator coil. Modes with $n \neq 0$ were checked with nomograph, Fig. 1.

Figure 3 shows that the calculated value of the resonance frequency pertaining to the $TE_{010}^{\infty\infty}$ mode is correct only for $L \rightarrow \infty$; and Fig. 4, that the one pertaining to the $TM_{110}^{0\infty}$ mode is correct only for $L \rightarrow 0$, as was explained in Sec. 2.2. Note that in Fig. 3 the abscissa is D/L, while in Fig. 4 it is L/D.

To check the validity of the theoretically obtained field configurations, cuts which do not interrupt the flow of dielectric displacement current were introduced, without affecting the resonance frequency. For the $TE_{010}^{\infty\infty}$ mode, the cylinder was composed of single disks of arbitrary thickness. No perceptible difference was observed as compared with a single piece of the same total length, although additional resonance frequencies could be detected (coupled circuits). Further proof was derived by varying the position of the exciting coil of the grid-dip-meter with respect to the test body. (Compare Figs. 5 and 6.) The loop with the variable condenser is the tuning circuit of the grid-dip-meter. Some magnetic flux lines are shown. If the magnetic lines of the grid-dip-meter circuit and of the cavity coincide, the cavity can be excited and these cases are marked "yes." In the positions shown marked "no," no excitation is possible.

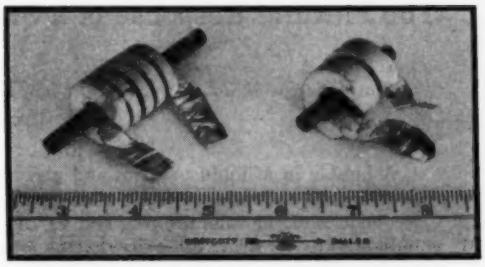


FIG. 7. Dielectric spiral antenna.

3.2 The Spiral Antenna as a Modified Application

At first sight the compactness of the cavities seems quite intriguing and leads easily to the idea of using them for miniaturization of cavities. However, at present the temperature coefficients of the high ϵ materials make them unfit for stable operation without temperature compensation. A change of several percent in frequency may be expected for a 50°C variation in temperature. Dielectrics having ϵ of the order of magnitude of 100 are much less temperature sensitive; their use seems much more practicable.

Another implication is that in the same ratio as the dimensions are minimized (the reduction is 50:1 for $\epsilon = 2500$), the impedance level is decreased. This, however, may be of advantage in circuits requiring low impedances.

Preliminary experiments were made using dielectric spirals for the excitation of magnetic dipole antennas in the vhf range (compare Fig. 7). The evolution of these spirals starts with cutting a wedge out of the $TE_{010}^{\infty\infty}$ mode cavity and connecting a transmission line to the silvered end surfaces. Several turns of dielectric are used in order to increase the input impedance; for instance, four dielectric turns resulted in 20 ohms for the first resonance point at 200 megacycles. At present the noise in the dielectric and the losses of magnetic materials in the vhf range prevent obtaining a high efficiency with this novel type of antenna. On the other hand, high-ϵ dielectric rods, which are "thin" in terms of the wavelength, can be equivalently substituted for copper rod dipole antennas. In this case the dielectric acts like a metal.

The Flexural Response of a Submerged Solid to a Pulsating Gas Bubble*

GEORGE CHERTOCK
David Taylor Model Basin, Washington, D. C., and Catholic University of America, Washington, D. C.
(Received October 21, 1952)

A bubble in a liquid pulsates and generates a changing pressure field. This reacts with a submerged flexible solid and induces elastic motions. Equations are derived which show how the vibration modes and frequencies of the solid are modified by the presence of the water, and how each mode is excited by the bubble pulsation. The generalized force for each mode is proportional to the volume acceleration of the bubble. The initial analysis presupposes incompressive flow, but it is shown that the results remain applicable if the duration of any compressive phase is small compared to the pulsation period of the bubble and the vibration period of the solid.

INTRODUCTION

THIS paper presents a theoretical analysis of the flexural response of a submerged solid to a pulsating gas bubble. The dynamics of such underwater bubbles has been extensively studied in the past, and there exists a large literature which describes the pulsation and migration of the bubble and the pressure field it generates.[1–4] However, this previous work has no reference to the interaction between the bubble and any flexible solid in its field.

The object of this investigation was to establish the physical mechanism responsible for the flexural motions of the solid, and to show quantitatively how these motions depend upon the elastic properties of the solid, and upon the size, position, and motion of the bubble. The method used in the analysis is to separate the motions of the solid and water into normal modes and then to show how each mode is coupled to the bubble pulsation.

NORMAL MODES OF THE SOLID

We shall consider only small vertical motions of an elongated solid for which the instantaneous displacement is uniform for all points on the same transverse section. Such motions may be generated by a vertical translation of the whole solid, or by a small rotation of the solid about a horizontal axis perpendicular to the length, and also by combined bending and shear vibrations in planes perpendicular to the length. For a solid which is long in comparison to its width or depth, there exists a set of normal modes of motion of this type, i.e., motions which the solid may have when it is moving freely, in vacuum, without applied external forces. These motions may be described by specifying the vertical displacement $\psi(x, t)$ of a longitudinal axis of the solid as a function of the time t and the distance x from one end:

$$\psi(x, t) \propto \psi_i(x)\cos\alpha_i t \quad i=0, 1, 2, \cdots. \tag{1}$$

The functions ψ_i describe the mode shapes. The index i which labels the various mode functions may conveniently be taken as equal to the number of nodes of ψ_i over the length of the solid. The frequencies α_i are the natural circular frequencies of the solid in vacuum. These mode shapes and frequencies completely specify the flexural properties of the solid and we shall take them as known functions. In practice, they may be determined by calculating the eigenfunctions and eigenvalues of some differential equation which governs the flexure, or perhaps by experiment.

In general, the mode functions are orthogonal over the length of the solid with $m(x)$, the mass per unit length, as a weighting function, so that

$$\int_0^l \psi_i\psi_j m dx = M_i\delta_{ij} \quad i, j=0, 1, 2\cdots, \tag{2}$$

where M_i is defined as the generalized mass associated with the ith mode. Also, we shall assume that these mode functions form a complete set, so that the most general dynamic configuration of the axis, even in forced motion can be expressed as a superposition

$$\psi(x, t)=\sum_i q_i(t)\psi_i(x), \tag{3}$$

where q_i are a set of generalized coordinates, each of which must satisfy an equation of the form

$$M_i(\ddot{q}_i+\alpha_i^2 q_i)=Q_i, \tag{4}$$

where Q_i is the generalized force on the solid. If this force is due solely to a distributed pressure acting on the surface of the solid, then

$$Q_i=-\int_S p\cos(n, k)\psi_i d\sigma, \tag{5}$$

where the integral is taken over the surface S, and (n, k) is the angle between the outward normal to the surface and the vertical.

These relations are all basic in vibration theory.

* This material is taken from some work supported by the Bureau of Ships and submitted to the Graduate School of the Catholic University of America in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

[1] C. Herring, "Theory of the pulsations of the gas bubble," NDRC Report C4-sr-30-027 (September, 1941).
[2] A. B. Arons, J. Acoust. Soc. Am. **20**, 277–82 (1948).
[3] M. S. Plesset, J. Appl. Phys. **16**, 277–83 (1949).
[4] B. Friedman, Commun. Pure Appl. Math. (NYU), 3 (1950).

MOTION OF THE WATER

Initially we shall assume that the flow generated by the pulsating bubble is an incompressive motion, but in a later section we will partially remove this restriction. Assume that there exists a velocity potential $\Phi(\xi, \eta, \zeta, t)$ which satisfies Laplace's equation at interior points ξ, η, ζ of the fluid, and whose gradient vanishes at infinity, and

$$-\partial\Phi/\partial n = v_g$$

on G, the surface of the bubble, where n is an outward normal and v_g is the local normal velocity of the bubble outward. Also,

$$-\partial\Phi/\partial n = \sum_i \dot{q}_i \psi_i \cos(n, k)$$

on S, the surface of the solid.

We construct Φ by superimposing several different potentials; thus

$$\Phi = \Phi_g(\xi, \eta, \zeta, t) + \sum_i \dot{q}_i(t)\varphi_i(\xi, \eta, \zeta), \tag{6}$$

where Φ_g describes the flow if the elastic solid were replaced by a rigid immovable surface and the bubble were constrained to maintain its original motion. Also φ_i describes the flow if the solid vibrates in a prescribed mode and the bubble were absent. Hence

$$-\partial\Phi_g/\partial n = 0; \quad -\partial\varphi_i/\partial n = \psi_i \cos(n, k) \text{ on } S, \tag{7}$$

$$-\partial\Phi_g/\partial n = v_g; \quad \partial\varphi_i/\partial n = 0 \text{ on } G. \tag{8}$$

This last condition makes φ_i a function of time as well as the space coordinates, because the surface G varies with time. However, we shall restrict the further analysis to cases where the bubble is sufficiently far from S so that its presence has negligible effect on defining the potential φ_i. In that case φ_i will be the same as the potential due to the prescribed vibration of S when the bubble is absent, and φ_i can be taken as a function of ξ, η, ζ only.

Now the pressure exerted on S by this flow is

$$\rho\frac{\partial\Phi}{\partial t} = \rho\frac{\partial\Phi_g}{\partial t} + \sum_j \rho\ddot{q}_j\varphi_j \tag{9}$$

and after substituting in (4) and (5), we have

$$M_i(\ddot{q}_i + \alpha_i^2 q_i) = -\int_S \rho\frac{\partial\Phi_g}{\partial t}\psi_i \cos(n, k)d\sigma - \sum_j \rho\frac{d^2q_j}{dt^2}\int_S \varphi_j\psi_i \cos(n, k)d\sigma \tag{10}$$

as the differential equation for the motion of the solid in response to the pulsating bubble. We can transform the first term on the right by applying Green's second identity to the two functions Φ_g and φ_i:

$$\int_S (\Phi_g\partial\varphi_i/\partial n - \varphi_i\partial\Phi_g/\partial n)d\sigma = -\int_G (\Phi_g\partial\varphi_i/\partial n - \varphi_i\partial\Phi_g/\partial n)d\sigma, \tag{11}$$

$$\int_S \Phi_g\psi_i \cos(n, k)d\sigma = \int_G (\Phi_g\partial\varphi_j/\partial n)d\sigma + \int_G \varphi_i v_g d\sigma. \tag{12}$$

Now if the bubble is far from S, then the bubble must be approximately spherical and v_g must be approximately uniform over G. Also the mean value of φ_i over G is the value which would prevail at the center of G if the bubble were removed. The presence of G adds a component to φ_i which is essentially a dipole term and which averages to zero over G. Hence

$$\int_G \varphi_i v_g d\sigma \simeq \varphi_{ig} dV/dt, \tag{13}$$

where V is the volume of the bubble and φ_{ig} is the value for φ_i at the position of the center of G if the bubble were removed.

Also we can transform the last term in (10) by defining an entrained mass of water,

$$L_{ij} = -\rho\int_S (\varphi_j\partial\varphi_i/\partial n)d\sigma = \rho\int_S \varphi_j\psi_i \cos(n, k)d\sigma = L_{ji}. \tag{14}$$

The symmetry properties of L_{ij} follow from Green's identity applied to φ_i and φ_j. Substituting (12), (13), and (14) into (10),

$$(M_i + L_{ii})(\ddot{q}_i + \omega_i^2 q_i) + \sum_{j\neq i} L_{ij}\ddot{q}_j = -\rho\varphi_{ig}\frac{d^2V}{dt^2}, \tag{15}$$

where

$$\omega_i^2 = \alpha_i^2 M_i/(M_i + L_{ii}). \tag{16}$$

It appears that each mode of motion behaves like a linear oscillator which is inertia coupled to each of the other modes. The generalized mass has been augmented by an entrained mass of water L_{ii} and this increased mass is also effective in decreasing the apparent natural frequency according to Eq. (16). The coefficients $L_{ij}(i \neq j)$ measure how much the inertia of the water couples together two modes which had been independent for vibrations in air. These coefficients may vanish if there is sufficient symmetry. For example, suppose the solid is symmetric about a vertical transverse plane through the mid-length. Then the functions ψ_i will be symmetrical or antisymmetrical about this midplane depending on whether i is even or odd. φ_i must have the same symmetry as ψ_i. Hence for such a solid, whenever i and j are of opposite parity, L_{ij} will vanish.

For the more general case of a solid without such symmetry, there exists a new set of coordinates, $q_i'(t)$,

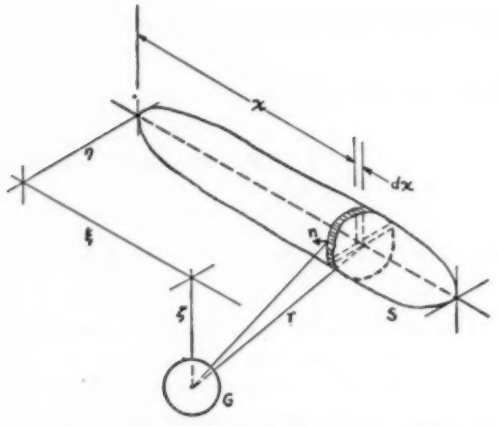


FIG. 1. Coordinate system for submerged solid and pulsating gas bubble.

and mode shapes $\psi_i'(x)$, such that in these new coordinates all modes are uncoupled. The transformation to these new variables is best carried out in matrix notation and this is done in the Appendix. The important point here is that in terms of these transformed variables the differential equation becomes

$$\ddot{q}_i'+\omega_i'^2 q_i'=-\rho\varphi_{i_g}'d^2V/dt^2. \tag{17}$$

This means that the solid may vibrate in water with a new set of mode shapes and frequencies, and that when this solid is exposed to a pulsating bubble, the generalized force for each mode of motion will be proportional to the volume acceleration of the bubble. Since we have neglected the influence of the solid in modifying the motion of the bubble, the volume acceleration to be used in Eq. (17) is that which would apply to the motion of a bubble pulsating in a free field. The equations for this free field motion are given in the literature. They may be solved numerically, or the main characteristics of the pulsation may be estimated from tabulated data.[1–4] We might mention that the presence of the bubble will modify the pressure field from the bubble and set up a pressure gradient at the bubble which will attract it to the solid, while the vibration of the solid will generate a fluctuating pressure at the bubble which will modify the pulsation period. However, both of these effects should be negligible if the bubble is farther than, say, one diameter away from the solid.

In the general case, the function φ_{i_g}' to be used in Eq. (17) must be determined by the solution of a complicated boundary value problem. However, we can derive an approximation for φ_{i_g}' which is useful over a wide range of particular cases. This approximation applies when (1) the distance from the bubble to a point on the surface is large compared to the cross-section dimensions of the solid, and (2) the solid is symmetric about a horizontal plane containing the longitudinal axis of the solid.

We start with the equation for φ_i' in terms of the values of the function and its gradient on S (see Fig. 1)

$$4\pi\varphi_{i_g}'=\int_S \varphi_i'\frac{\partial}{\partial n}\left(\frac{1}{r_g}\right)d\sigma-\int_S\frac{1}{r_g}\frac{\partial\varphi_i'}{\partial n}d\sigma. \tag{18}$$

Now cut the solid into disks of width dx by transverse planes perpendicular to its axis, and let ds be an element of the perimeter around the disk at x. Then

$$4\pi\varphi_{i_g}'=\int_0^l\oint\varphi_i'\frac{\partial}{\partial n}\left(\frac{1}{r_g}\right)dsdx+\int_0^l\dot{\psi}_i'\oint\frac{1}{r_g}\cos(n,k)dsdx. \tag{19}$$

Consider the special case for which r_g is much larger than the cross section of the disk, and let r be the distance from G to the axis. Then

$$\oint\varphi_i'\frac{\partial}{\partial n}\left(\frac{1}{r_g}\right)ds=-\oint\varphi_i'\frac{\cos(n,r_g)}{r_g^2}ds\simeq-\oint\varphi_i'\frac{\cos(n,r)}{r^2}ds \tag{20}$$

or, in terms of the direction cosines of r and n,

$$\oint\varphi_i'\frac{\partial}{\partial n}\left(\frac{1}{r_g}\right)ds=-\frac{x-\xi}{r^3}\oint\varphi_i'n_xds-\frac{\eta}{r^3}\oint\varphi_i'n_yds-\frac{\zeta}{r^3}\oint\varphi_i'n_zds. \tag{21}$$

Because of the assumed symmetry about the horizontal mid-plane, n_x and n_y must both be symmetric about this plane, while n_z must be antisymmetric. Hence the first two integrals on the right must vanish. The integrand of the third term can be transformed by Eq. (58) of the Appendix, so that (21) now reduces to

$$\oint\varphi_i'\frac{\partial}{\partial n}\left(\frac{1}{r_g}\right)ds=\frac{m}{\rho}\frac{\zeta}{r^3}(\psi_i'-\psi_i''), \tag{22}$$

where ψ_i'' is a particular combination of the ψ_i' function as described in the Appendix. Now returning to the second term in (19)

$$\oint\frac{1}{r_g}\cos(n,k)dsdx=\int\operatorname{div}\frac{k}{r_g}dAdx\simeq-\frac{\zeta}{r^3}Adx, \tag{23}$$

where A is the section area of the disk, and the left side may be considered as an integral over the entire surface of the disk. The second integral follows from Gauss' theorem. Note that ζ is the depth of G below the axis of S. Substituting (22) and (23) into (19),

$$4\pi\varphi_{i_g}'=-\frac{\zeta}{\rho}\int_0^l\left(\frac{m}{r^3}\psi_i''+\frac{A\rho-m}{r^3}\psi_i'\right)dx. \tag{24}$$

This last result may be considered as a generalization of an analysis given by Lamb for the effect at a distance due to the rigid translation of a solid in a liquid.[5]

Finally substituting this value for $\varphi_{i\sigma}'$ into Eq. (17),

$$\ddot{q}_i'+\omega_i'^2 q_i'=\left[\frac{\zeta}{4\pi}\int_0^l\left(\frac{m}{r^3}\psi_i''+\frac{A\rho-m}{r^3}\psi_i'\right)dx\right]\frac{d^2V}{dt^2}. \quad (25)$$

In this equation, all the parameters may be determined experimentally by observing the free vibrations of the solid in water and the gas bubble in a free field.

Proportional Solid

The simplest version of this analysis occurs in the special case where the density of the solid is uniform and equal to $\rho=m/A$ and where the solid is "proportional." We define a proportional solid as one for which the

$$\oint \varphi_i' \cos(n,k)ds=a_i m\psi_i', \quad (26)$$

where a_i is a constant, independent of x. Multiply this equation by ψ_j and integrate over the length of the solid.

$$\int_0^l \psi_j \oint \varphi_i \cos(n,k)dsdx=a_i\int_0^l m\psi_i\psi_j dx, \quad (27)$$

$$L_{ij}=a_i\rho M_i\delta_{ij}. \quad (28)$$

Hence a proportional solid is one for which the cross coupling coefficients of (15) vanish, and the transformation to new coordinates is trivial, i.e.,

$$q_i'=q_i(M_i+L_{ii})^{\frac{1}{2}};\quad \psi_i'=\psi_i(M_i+L_{ii})^{-\frac{1}{2}},$$
$$\omega_i'=\omega_i=\alpha_i(M_i/M_i+L_{ii})^{\frac{1}{2}};$$
$$\psi_i''=\psi_i(M_i+L_{ii})^{\frac{1}{2}}/M_i, \quad (29)$$

and substituting in (25),

$$\ddot{q}_i+\omega_i^2 q_i=\left[\frac{\rho\zeta}{4\pi M_i}\int_0^l\frac{A\psi_i}{r^3}dx\right]\frac{d^2V}{dt^2}. \quad (30)$$

The entrained mass has disappeared from the equation except for its effect on the natural frequency. This notion of a proportional solid is useful in practice where the mode shapes in air and water are the same experimentally.

A simple consequence of this applies to the vertical translation of the solid. This may be considered a mode for which $\psi_i=\psi_0$, a constant, and $\omega_0=0$. The terms in Eq. (30) now no longer have any dependence on the elastic properties of the solid, and they may equally well apply to a body of water with the same shape as the surface of the solid. Hence the vertical motion of the center of gravity of the proportional solid is the same as the vertical motion of the centroid of the displaced water in the absence of the solid. It must be understood that we are talking about displacements which are small compared to the dimensions of the solid.

More generally, if we consider a very rigid solid of arbitrary density, then all modes of motion should be negligible except those which describe vertical translation and rotation, and these modes can not be appreciably coupled to the flexure modes. Assume sufficient symmetry to eliminate coupling between translation and rotation, and denoting the translation mode by the subscript 0, then effectively the body acts as a proportional solid for this mode, and

$$\psi_0=1;\quad q_0'=q_0(M_0+L_{00})^{\frac{1}{2}},$$
$$\psi_0''=(M_0+L_{00})^{\frac{1}{2}}/M_0;\quad \psi_0'=(M_0+L_{00})^{-\frac{1}{2}}, \quad (31)$$

and substituting in (25),

$$\ddot{q}_0=\left[\frac{\zeta}{4\pi}\int_0^l\frac{1}{r^3}\left(\frac{m}{M_0}+\frac{A\rho-m}{M_0+L_{00}}\right)dx\right]\frac{d^2V}{dt^2}, \quad (32)$$

and consider only the case for which $r\gg l$

$$\ddot{q}_0=\frac{\zeta}{4\pi r^3}\frac{M_w+L_{00}}{M_0+L_{00}}\frac{d^2V}{dt^2}, \quad (33)$$

where M_w is the mass of the displaced water. Hence when everything starts from rest, the displacement is simply proportional to the bubble volume.

FLEXURAL MOTIONS OF A FLOATING SOLID

The problem of the flexural reactions of a floating solid to a pulsating bubble can be treated by the same methods described for the submerged solid. An added complication is that besides the other boundary conditions, the velocity potential must vanish at the free surface. This condition is necessary in order to approximately satisfy the condition that the pressure at the free surface must always equal atmospheric pressure.

This condition may be satisfied by the usual device of constructing an image space which is the reflection of the fluid space in the free surface, and where the potential at any point is the negative of the potential at the corresponding point in the fluid. Hence if S' is the image of S and G' is the image of the bubble G (see Fig. 2), then Φ_σ and φ_i must satisfy the additional boundary conditions

$$-\partial\Phi_\sigma/\partial n=v_\sigma \text{ on } G', \quad (34)$$

$$-\partial\Phi_\sigma/\partial n=0 \text{ on } S', \quad (35)$$

$$-\partial\varphi_i/\partial n=\psi_i\cos(n,k) \text{ on } S', \quad (36)$$

where v_σ on G' equals $-v_\sigma$ for the image point on G, and $\cos(n,k)$ on S' equals $-\cos(n,k)$ for the image point on S. Then the same analysis leads to the same differential equation as (17):

$$\ddot{q}_i'+\omega_i'^2 q_i'=-\rho\varphi_{i\sigma}'\frac{d^2V}{dt^2}=-\rho\int_S\frac{\partial\Phi_\sigma}{\partial t}\psi_i'\cos(n,k)d\sigma, \quad (37)$$

[5] H. Lamb, Hydrodynamics (Cambridge University Press, Cambridge, 1932), sixth edition, art. 121a.

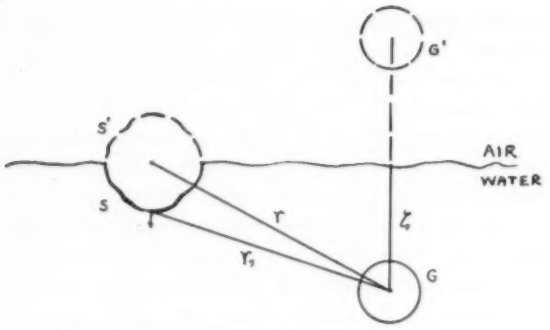


FIG. 2. Floating solid and pulsating gas bubble.

and

$$L_{ii'}=\rho\int_S \varphi_i'\psi_i'\cos(n,k)d\sigma. \tag{38}$$

The integration is over the underwater area only. However, in order to evaluate φ_{i_g}' by Green's function

$$4\pi\varphi_{i_g}'=\int_{S+S'}\left[\varphi_i'\frac{\partial}{\partial n}\left(\frac{1}{r_g}\right)-\frac{1}{r_g}\frac{\partial\varphi_i'}{\partial n}\right]d\sigma, \tag{39}$$

the integration must be made over S and S'. This leads to a factor two in the equation for φ_{i_g}' under conditions analogous to (24); i.e.,

$$-4\pi\varphi_{i_g}'=\frac{2\zeta}{\rho}\int_0^l\frac{m}{r^3}\psi_i''dx+\frac{2\zeta}{\rho}\int_0^l\frac{A\rho-m}{r^3}\psi_i'dx, \tag{40}$$

where A is now the area of the underwater section of the solid at a distance x from one end.

EFFECT OF AN INITIAL COMPRESSIVE WAVE

The foregoing was all based on the assumption that the flow which is generated by the bubble pulsation is an incompressive motion. This is an assumption common in bubble theory and is known to be valid during a large fraction of the bubble oscillation period when the radial velocity of the bubble surface is small compared to the velocity of sound. However, in the initial stages of its motion, when, for example, a bubble is created, a compressive wave may be emitted and the velocity potential may differ markedly from a harmonic function. Hence we shall assume that this flow satisfies the wave equation

$$c^2\Delta\Phi_g=\partial^2\Phi_g/\partial t^2, \tag{41}$$

and shall determine under what conditions the incompressive analysis remains valid. We still assume that the φ_i functions are harmonic functions.

Then the previous analysis remains applicable up to Eq. (11) where the application of Green's identity results in

$$\int_S\Phi_g\psi_i\cos(n,k)d\sigma=\int_G\varphi_iv_gd\sigma-\int\varphi_i\Delta\Phi_gd\tau, \tag{42}$$

where the last integral is taken over the whole volume of water. This integral can be written

$$\int\varphi_i\Delta\Phi_gd\tau=\int\frac{\varphi_i}{c^2}\frac{\partial^2\Phi_g}{\partial t^2}d\tau=\int\frac{\varphi_i}{\rho c^2}\frac{\partial p}{\partial t}d\tau, \tag{43}$$

where p is the local excess pressure due to the bubble. Hence substituting (42) and (43) into (41),

$$M_i(\ddot{q}_i+\alpha_i^2q_i)+\sum_jL_{ij}\ddot{q}_j=-\rho\varphi_{i_g}\frac{d^2V}{dt^2}+\int\frac{\varphi_i}{c^2}\frac{\partial^2p}{\partial t^2}d\tau. \tag{44}$$

Integrate twice with respect to time starting with time zero when everything is at rest and the bubble is created:

$$M_i\left[q_i+\alpha_i^2\int_0^t\int_0^t q_idtdt\right]+\sum_jL_{ij}q_j$$

$$=-\rho\varphi_{i_g}V+\int(\varphi_ip/c^2)d\tau. \tag{45}$$

Consider the integrand of the last term. φ_i is a function which decreases faster than (ζ/r^3) from the target, also it averages to zero about the target. $p(\xi,\eta,\zeta,t)$ is initially very large at the bubble, but this phase travels off into the water, the amplitude decaying as r^{-1}. Furthermore, it is known that p becomes negative at the bubble after eleven percent of the bubble period, and this negative phase travels out into the water. Hence it is clear that shortly after the creation of the bubble, a time t will exist when the last term in (45) is negligible or even zero compared to the first term on the right. If also this time is small compared to the natural period of oscillation in the ith mode, i.e., if $\alpha t\ll 2\pi$, then the integral term on the left side of this equation must be negligible compared to the first term. Hence at this time the displacement q_i as calculated from Eq. (45) will be the same whether or not the compressibility term is included.

This means that the incompressive theory is adequate to explain the flexural vibrations of the solid even though the solid is initially exposed to a compressive wave, provided that the period of oscillation is long compared to the duration of the compressive wave. On the other hand, for high frequency motions of the solid the generalized forces are not related to d^2V/dt^2 in any simple way, and the theory is not applicable.

APPENDIX

We derive here some properties of the mode functions and natural frequencies for the solid in water. Let $\mathbf{q}$, $\boldsymbol{\psi}$, $\boldsymbol{\phi}$, $\mathbf{L}$, $\mathbf{M}$, $\boldsymbol{\alpha}$ denote the matrices with elements q_i, ψ_i, φ_i, L_{ij}, M_i, α_i, respectively. $\mathbf{q}$, $\boldsymbol{\psi}$, and $\boldsymbol{\phi}$ are column matrices; $\mathbf{M}$ and $\boldsymbol{\alpha}$ are diagonal matrices, and $\mathbf{L}$ is symmetrical. We will denote the reciprocal matrix by $\mathbf{L}^{-1}$, and the transpose matrix by $\tilde{\mathbf{L}}$. Then the

matrix analog of Eq. (15) becomes

$$(\mathbf{M}+\mathbf{L})d^2\mathbf{q}/dt^2+\mathbf{M}\alpha^2\mathbf{q}=-\rho\frac{d^2V}{dt^2}\boldsymbol{\phi}_\sigma, \tag{46}$$

and after some algebra,

$$\mathbf{M}^{\frac{1}{2}}\alpha d^2\mathbf{q}/dt^2+[\mathbf{M}^{\frac{1}{2}}\alpha(\mathbf{M}+\mathbf{L})^{-1}\mathbf{M}^{\frac{1}{2}}\alpha]\mathbf{M}^{\frac{1}{2}}\alpha\mathbf{q}$$
$$=-\rho[\mathbf{M}^{\frac{1}{2}}\alpha(\mathbf{M}+\mathbf{L})^{-1}\mathbf{M}^{\frac{1}{2}}\alpha]\mathbf{M}^{-\frac{1}{2}}\alpha^{-1}\boldsymbol{\phi}_\sigma d^2V/dt^2. \tag{47}$$

Now diagonalize the matrix in brackets by means of an orthogonality transformation

$$\mathbf{M}^{\frac{1}{2}}\alpha(\mathbf{M}+\mathbf{L})^{-1}\mathbf{M}^{\frac{1}{2}}\alpha=\mathbf{R}^{-1}\boldsymbol{\Omega}^2\mathbf{R}, \tag{48}$$

where $\boldsymbol{\Omega}^2$ is a diagonal matrix whose elements ω_i^2 are the characteristic roots of the matrix in brackets, and $\mathbf{R}$ is an orthogonal matrix which is compounded of the characteristic solutions of the matrix in brackets. Hence

$$\boldsymbol{\Omega}^{-1}\mathbf{R}\mathbf{M}^{\frac{1}{2}}\alpha d^2\mathbf{q}/dt^2+\boldsymbol{\Omega}^2(\boldsymbol{\Omega}^{-1}\mathbf{R}\mathbf{M}^{\frac{1}{2}}\alpha)\mathbf{q}$$
$$=-\rho(\boldsymbol{\Omega}\mathbf{R}\mathbf{M}^{-\frac{1}{2}}\alpha^{-1})\boldsymbol{\phi}_\sigma d^2V/dt^2, \tag{49}$$

or

$$d^2\mathbf{q}'/dt^2+\boldsymbol{\Omega}^2\mathbf{q}'=-\rho(d^2V/dt^2)\boldsymbol{\phi}_\sigma', \tag{50}$$

where we have defined a transformed set of normal coordinates, mode shapes, and potentials,

$$\mathbf{q}'=\mathbf{A}\mathbf{q};\quad \boldsymbol{\phi}'=\tilde{\mathbf{A}}^{-1}\boldsymbol{\phi};\quad \boldsymbol{\psi}'=\tilde{\mathbf{A}}^{-1}\boldsymbol{\psi}, \tag{51}$$

where $\mathbf{A}$ is the transformation matrix $\boldsymbol{\Omega}^{-1}\mathbf{R}\alpha\mathbf{M}^{\frac{1}{2}}$. Also we can define a new set of generalized masses

$$\mathbf{M}'=\int_0^l \boldsymbol{\psi}'\tilde{\boldsymbol{\psi}}'m\,dx=\tilde{\mathbf{A}}^{-1}\mathbf{M}\mathbf{A}^{-1}, \tag{52}$$

$$\mathbf{L}'=\int_S \boldsymbol{\phi}'\tilde{\boldsymbol{\psi}}'\rho\cos(n,k)\,d\sigma=\tilde{\mathbf{A}}^{-1}\mathbf{L}\mathbf{A}^{-1}, \tag{53}$$

where $\mathbf{M}'+\mathbf{L}'=\mathbf{I}$ the unit matrix. Equation (50) is the justification for Eq. (17).

Now we inquire into the relation between the new mode functions $\psi_i'(x)$ and the values for φ_i' on S. First we define a series of functions $\bar{\varphi}_i'$ as an integral of φ_i' around the perimeter s of a section of the solid at the distance x from one end, i.e.,

$$\bar{\varphi}_i'(x)=\oint \varphi_i'\cos(n,k)\,ds. \tag{54}$$

We expand $\bar{\varphi}_i'$ in terms of the set of functions $m\psi_i'$, or in matrix notation

$$\bar{\boldsymbol{\phi}}'=m\mathbf{B}\boldsymbol{\psi}', \tag{55}$$

where $\mathbf{B}$ is an undetermined matrix with constant coefficients. In order to determine $\mathbf{B}$, postmultiply both sides of (55) by $\tilde{\boldsymbol{\psi}}'$ and integrate with respect to x:

$$\int_0^l \bar{\boldsymbol{\phi}}'\tilde{\boldsymbol{\psi}}'dx=\mathbf{B}\int_0^l \boldsymbol{\psi}'\tilde{\boldsymbol{\psi}}'m\,dx \tag{56}$$

or

$$\mathbf{B}=\rho^{-1}\mathbf{L}'\mathbf{M}'^{-1}, \tag{57}$$

and substituting in (55),

$$\bar{\boldsymbol{\phi}}'=(\mathbf{M}'^{-1}\boldsymbol{\psi}'-\boldsymbol{\psi}')m/\rho=(\boldsymbol{\psi}''-\boldsymbol{\psi}')m/\rho, \tag{58}$$

where $\boldsymbol{\psi}''=\mathbf{M}'^{-1}\boldsymbol{\psi}'$ represents a column matrix of mode functions which are a particular linear combination of the ψ_i' functions.

ACKNOWLEDGMENT

The author is indebted to Professor K. F. Herzfeld of the Catholic University of America and to Dr. E. H. Kennard and Dr. W. J. Sette of the David Taylor Model Basin for encouragement and criticism.

Forced Oscillations with Nonlinear Restoring Force*

CHIHIRO HAYASHI
Department of Electrical Engineering, University of Kyoto, Kyoto, Japan
(Received May 28, 1952)

Forced oscillations with nonlinear restoring force are studied in transient states as well as in steady states. The original differential equation characterized by a nonlinear term is transformed under certain restrictions to the following differential equation of the first order:

$$dy/dx = Y(x, y)/X(x, y).$$

Following Poincaré and Bendixson, the singularities and the integral curves of the above equation are discussed, the former being correlated with the periodic states of oscillations, the latter with the transient states of oscillations. The stability of the periodic solutions is determined in accordance with that of the singular points, *viz.*, according to the roots of the characteristic equation. The integral curves yield the relationship between the given initial conditions and the periodic solutions. Thus, once the initial conditions are prescribed, we can foresee the final periodic states which are started with those conditions.

With this method of investigation, we have first studied the harmonic oscillations presented in Part I and then the subharmonic oscillation of order $\frac{1}{3}$ given in Part II. In both cases the theoretical results are compared with the experimental measurements carried out for an electric circuit containing a saturable iron core, and the satisfactory agreement is found between them.

INTRODUCTION

A DIFFERENTIAL equation of the form

$$\frac{d^2v}{d\tau^2}+k\frac{dv}{d\tau}+f(v)=B\cos\tau, \tag{1.1}$$

in which $f(v)$ is a nonlinear function of v, occurs in several different kinds of physical problems. As will be shown later, the case of alternating current circuits containing saturable iron core inductances is an example which gives rise to Eq. (1.1).

From the theory of differential equations, it is known that the above Eq. (1.1) possesses such solutions $v(\tau)$ that are uniquely determined once the values of $v(0)=v_{\tau=0}$ and $(dv/d\tau)_{\tau=0}$, the initial conditions, are prescribed. In the case of linear oscillations where $f(v)$ is proportional to v, there is one and only one periodic solution after some transients have damped out. It is, however, the distinctive character of nonlinear oscillations that various types of periodic solutions of Eq. (1.1) may exist corresponding to different values of the initial conditions prescribed.

In the following lines we shall concentrate our attention to the stability of periodic solutions and, in particular, to the transient states of oscillations which give rise to the periodic steady states according to the initial conditions prescribed.

PART I. HARMONIC OSCILLATIONS

1. Periodic Solutions and Their Stability

We shall first deal with the harmonic oscillations, namely, the periodic solutions $v(\tau)$ of Eq. (1.1) in which the period is the same as the period 2π of the external force $B\cos\tau$. Subharmonic oscillations, in which the solution $v(\tau)$ has as its least period an integral multiple (different from unity) of the period of the external force, will be treated later in Part II.

Now, to discuss the problem in the concrete, let us choose for $f(v)$ in Eq. (1.1) the following simple but exemplary nonlinear relation:†

$$f(v)=v^3; \tag{1.2}$$

then Eq. (1.1) becomes

$$\frac{d^2v}{d\tau^2}+k\frac{dv}{d\tau}+v^3=B\cos\tau. \tag{1.3}$$

In the case of harmonic oscillations in which the oscillation of fundamental frequency predominates over the higher harmonics, this equation admits a solution of the form‡

$$v(\tau)=x(\tau)\sin\tau+y(\tau)\cos\tau, \tag{1.4}$$

where the amplitudes $x(\tau)$ and $y(\tau)$ are both functions

* Submitted to the Faculty of Engineering of Kyoto University in partial fulfillment of the requirements for the degree of Doctor of Science, October, 1951.

† When the linear term is retained, i.e., $f(v)=c_1v+c_3v^3$, c_1, c_3 being constants, the investigation may also be carried out in the same manner. See further the footnote in Sec. 4.

‡ It has been pointed out by E. Trefftz [Math. Ann. **95**, 307 (1925)] that if a solution of (1.3) is stable, it must ultimately lead to a periodic solution in which the period is equal to the period 2π of the external force (the harmonic oscillation), or as its least period equal to an integral multiple of 2π (the subharmonic oscillation). In the case when we deal with the harmonic oscillation, the Fourier series development of the periodic solution may be started with $x\sin\tau+y\cos\tau$. The higher harmonic terms are omitted in Eq. (1.4). This assumption may, however, cause no serious error, because (1) from the physical point of view the equation of the system should expediently be set up with respect to a quantity in which the higher harmonic content is as low as possible, and, as one will see later in Sec. 4, the magnetic flux in the iron core is represented by v, which contains the less harmonic content compared to the current in the coil; and (2), when the external force B is not so large, the coefficients of the Fourier series decrease rather rapidly with the increase of the harmonic order. We shall, in the near future, report on the higher harmonic oscillation in the case when B is large.

of τ, but are ultimately reduced to constants after the transient states have elapsed.

Substituting (1.4) into (1.3) and separately equating the terms containing $\sin\tau$ and $\cos\tau$ to zero, we obtain

$$\left.\begin{aligned} \frac{dx}{d\tau}&=\tfrac{1}{2}[B-kx+y-\tfrac{3}{4}r^2y]\equiv X(x,y),\\ \frac{dy}{d\tau}&=\tfrac{1}{2}[-x-ky+\tfrac{3}{4}r^2x]\equiv Y(x,y),\\ \text{where}\quad r^2&=x^2+y^2, \end{aligned}\right\} \tag{1.5}$$

under the restrictions that (1) the amplitudes $x(\tau)$, $y(\tau)$ vary only slowly with the time τ, *viz.*,

$$dx/d\tau \ll x, \qquad dy/d\tau \ll y,$$
$$d^2x/d\tau^2 \ll dx/d\tau, \quad d^2y/d\tau^2 \ll dy/d\tau,$$

so that $d^2x/d\tau^2$ and $d^2y/d\tau^2$ may be neglected, and that (2) the damping coefficient k is a relatively small quantity.

The results derived from Eqs. (1.5) may therefore be not applicable to the case in which the higher harmonic oscillations are sustained. But, as far as we deal with the harmonic oscillations, Eqs. (1.5) may be assumed to be legitimate.

In the steady states where the amplitudes $x(\tau)$, $y(\tau)$ in Eq. (1.4) are constant, we have

$$\left.\begin{aligned} dx/d\tau&=X(x,y)=0,\\ dy/d\tau&=Y(x,y)=0. \end{aligned}\right\} \tag{1.6}$$

From Eqs. (1.5) and (1.6) the amplitude r_0 of the periodic solution $v(\tau)$ is given by

$$\left.\begin{aligned} B^2&=r_0^2(A^2+k^2),\\ \text{with}\quad A&=\tfrac{3}{4}r_0^2-1, \end{aligned}\right\} \tag{1.7}$$

and the components x_0, y_0 of r_0 are found as

$$\left.\begin{aligned} &Bx_0=kr_0^2, \quad Ax_0=ky_0,\\ \text{or}\quad &x_0^2=\frac{r_0^2}{1+(A/k)^2}, \quad y_0^2=\frac{r_0^2}{1+(k/A)^2}. \end{aligned}\right\} \tag{1.8}$$

The circumstances under which this condition obtains can be determined by a further stability investigation. To effect this we shall consider small variations ξ and η from the solutions x_0 and y_0 as given by Eqs. (1.8), and determine whether these deviations approach zero or not with increase of the time τ.

From Eqs. (1.5) we have

$$\left.\begin{aligned} \frac{d\xi}{d\tau}&=\frac{\partial X}{\partial x}\xi+\frac{\partial X}{\partial y}\eta,\\ \frac{d\eta}{d\tau}&=\frac{\partial Y}{\partial x}\xi+\frac{\partial Y}{\partial y}\eta. \end{aligned}\right\} \tag{1.9}$$

The solutions of these simultaneous equations have the form $e^{\lambda\tau}$, and λ is determined by a characteristic equation:

$$\begin{vmatrix} \dfrac{\partial X}{\partial x}-\lambda & \dfrac{\partial X}{\partial y} \\ \dfrac{\partial Y}{\partial x} & \dfrac{\partial Y}{\partial y}-\lambda \end{vmatrix}=0. \tag{1.10}$$

The variations ξ and η approach zero with increase of the time τ, provided that the real part of λ is negative, and the corresponding periodic solutions determined from Eqs. (1.4) and (1.8) are stable. This stability condition§ is given by Hurwitz's criterion[1] as

$$\left.\begin{aligned} &-\frac{\partial X}{\partial x}-\frac{\partial Y}{\partial y}>0,\\ \text{and}\quad &\frac{\partial X}{\partial x}\frac{\partial Y}{\partial y}-\frac{\partial Y}{\partial x}\frac{\partial X}{\partial y}>0. \end{aligned}\right\} \tag{1.11}$$

Substituting (1.5) and (1.7) into (1.11), we obtain

$$\left.\begin{aligned} &k>0,\\ \text{and}\quad &\frac{27}{16}r_0^4-3r_0^2+k^2+1>0 \quad \text{or} \quad \frac{dB^2}{dr_0^2}>0. \end{aligned}\right\} \tag{1.12}$$

Evidently the first condition is fulfilled from the outset, since we are here concerned with the positive damping. The second condition shows that the periodic solution is stable under such circumstances that the amplitude r_0 increases with increasing external force B. This is quite a plausible condition.

2. Analysis of the Nonlinear Oscillations by Means of Integral Curves

As mentioned already, our object is to study, in particular, the transient solutions of Eq. (1.3), which, with the lapse of time, yield ultimately to the periodic solutions. For this purpose it is useful to investigate, following Poincaré[2] and Bendixson,[3] the integral curves

§ This condition may fail if the higher harmonic oscillations are predominantly superposed on the harmonic oscillation with negative damping. A more rigorous criterion for the stability will be given in a later paper.

[1] A. Hurwitz, Math. Ann. **46**, 273 (1895).

[2] H. Poincaré, *Oeuvres* (Gauthier-Villars, Paris, 1928), Vol. I. pp. 1–222.

[3] I. Bendixson, Acta Math. **24**, 1 (1901).

of the equation

$$\frac{dy}{dx}=\frac{Y(x, y)}{X(x, y)}, \tag{2.1}$$

which has been derived from Eqs. (1.5).

Since the time τ does not occur explicitly in this equation, we can draw the integral curves in the x, y plane with the aid of the isocline method or otherwise.‖ As mentioned in the preceding section, periodic solutions are related with $x(\tau)=$ constant, $y(\tau)=$ constant of Eqs. (1.5), and therefore with the singular points of Eq. (2.1), i.e., the points at which $X(x, y)$ and $Y(x, y)$ vanish simultaneously.

Now suppose that we fix a point $x(0), y(0)$ in the x, y plane as the initial condition. Then the point $x(\tau), y(\tau)$ moves with the lapse of time τ on the integral curve which passes through the initial point $x(0), y(0)$ and gets finally to the stable singular points.¶ Hence the transient solutions are correlated with the integral curves of Eq. (2.1), and the time variation of $v(\tau)$ is obtained by the following line integral as derived from Eqs. (1.5):

$$\tau=\int\frac{ds}{[X^2(x, y)+Y^2(x, y)]^{\frac{1}{2}}}, \qquad ds=[(dx)^2+(dy)^2]^{\frac{1}{2}}, \tag{2.2}$$

where ds is the line element on the integral curve.

We turn, for a time being, to a discussion of the character of the integral curves in the neighborhood of the singular points of Eq. (2.1). For the singular points, the roots of the characteristic Eq. (1.10) are given by Eqs. (1.5) as

$$\lambda_{1,2}=\frac{a_1+b_2\pm[(a_1-b_2)^2+4a_2b_1]^{\frac{1}{2}}}{2},$$

with

$$\left.\begin{aligned} a_1&=\left(\frac{\partial X}{\partial x}\right)_{\substack{x=x_0,\\ y=y_0}}=\tfrac{1}{2}[-k-\tfrac{3}{2}x_0y_0],\\ a_2&=\left(\frac{\partial X}{\partial y}\right)_{\substack{x=x_0,\\ y=y_0}}=\tfrac{1}{2}[1-\tfrac{3}{4}(x_0^2+3y_0^2)],\\ b_1&=\left(\frac{\partial Y}{\partial x}\right)_{\substack{x=x_0,\\ y=y_0}}=\tfrac{1}{2}[-1+\tfrac{3}{4}(3x_0^2+y_0^2)],\\ b_2&=\left(\frac{\partial Y}{\partial y}\right)_{\substack{x=x_0,\\ y=y_0}}=\tfrac{1}{2}[-k+\tfrac{3}{2}x_0y_0]. \end{aligned}\right\} \tag{2.3}$$

Poincaré[4] has classified the types of singular points according to the character of the integral curves near the singular points, *viz.*, according to the nature of the characteristic roots λ, as follows:

(1) The singularity is a nodal point, if the characteristic roots are both real numbers with same sign, so that

$$\left.\begin{aligned} (a_1-b_2)^2+4a_2b_1&\geqq 0,\\ a_1b_2-a_2b_1&>0. \end{aligned}\right\} \tag{2.4}$$

(2) The singularity is a saddle point, if the two roots are real numbers with opposite signs, so that

$$\left.\begin{aligned} (a_1-b_2)^2+4a_2b_1&>0,\\ a_1b_2-a_2b_1&<0. \end{aligned}\right\} \tag{2.5}$$

(3) The singularity is a spiral point, if the two roots are conjugate complex to each other, so that

$$(a_1-b_2)^2+4a_2b_1<0. \tag{2.6}$$

If, in particular, the two roots are both purely imaginary so that $a_1+b_2=0$, the singularity is either a center or a spiral.**

We also infer a singularity to be stable or unstable according as a point on any integral curve moves into the said singularity or not with increasing τ, i.e., according as the real part of λ is negative or positive.

Bearing the foregoing discussion in mind, we turn now to our present case of harmonic oscillations. The relation between B and r_0^2 as given by Eqs. (1.7) is shown in Fig. 1 for several values of k.

In accordance with the foregoing analysis, we shall classify these periodic states of equilibrium†† according to the types of the singularities. The boundary line between nodes and saddles is determined by Eqs. (2.4) and (2.5) as

$$a_1b_2-a_2b_1=0,$$

which, upon substituting Eqs. (2.3), leads to

$$\frac{27}{16}r_0^4-3r_0^2+k^2+1=0. \tag{2.7}$$

This equation is identical with $dB^2/dr_0^2=0$ [see Eqs. (1.12)], and the region of the saddle points is represented with shades where dB^2/dr_0^2 is negative. Since one of the roots λ is always positive, the periodic states in this region are unstable. This agrees with the result obtained in the preceding section.

Similarly, the boundary lines between nodes and

‖ Another form of graphical solution has been given by Liénard, Rev. Gén. de l'Élec. 23, 901, 946 (1928).

¶ According to Bendixson (see reference 3, p. 78), it will be verified without difficulty that the integral curves of Eq. (2.1) have no limit cycle provided that the original differential equation is of the form (1.1).

[4] See reference 2, p. 14.

** Poincaré (reference 2, p. 95) has shown the rigorous criterion for distinguishing between a center and a spiral in the case when $(a_1-b_2)^2+4a_2b_1<0$ and $a_1+b_2=0$.

†† These periodic states of equilibrium are not necessarily sustained, and they are capable for lasting out so long as they are correlated with the stable singularities.

spirals are determined by Eqs. (2.4) and (2.6) as

$$(a_1-b_2)^2+4a_2b_1=0,$$

which, upon substituting Eqs. (2.3), leads to

$$\frac{27}{16}r_0^4-3r_0^2+1=0, \quad \text{i.e.,} \quad r_0^2=\frac{4}{3}, \frac{4}{9}. \tag{2.8}$$

These boundaries are shown by dotted lines in Fig. 1. Since $a_1+b_2=-k$ by Eqs. (2.3), it is readily seen that the periodic states of equilibrium in the regions of nodes and spirals are evidently stable.

3. Geometrical Discussion of Integral Curves in Special Cases

In the preceding section we have briefly referred to the transient solutions which are correlated with the integral curves of Eq. (2.1). It is, however, useful and illuminating to consider the integral curves for some typical cases. The special cases we have in mind are those specified by the following values of k and B in Eq. (1.3), namely,

Case 1. $k=0.2$, $B=0.3$,

and

Case 2. $k=0.7$, $B=0.75$.

As observed in Fig. 1, there are three different states of equilibrium in the first Case 1, whereas we have only one periodic solution in the second Case 2. The integral curves for these two cases are shown in Figs. 2 and 3, respectively. The singularities are determined by Eqs. (1.6) and summed up in the following Table I.

The integral curves have been drawn with the aid of the isoclines as represented by dotted line curves, on which the numbers show the values of dy/dx for respective isoclines. By Eqs. (1.5), a point $x(\tau)$, $y(\tau)$ moves, with the lapse of time τ, on the integral curve as indicated by arrows and gets finally to the stable singular point.

In Fig. 2 we have three singularities; two of them, i.e., (1) and (3), are stable, and the corresponding periodic oscillations are sustained. The remaining singularity (2) is a saddle point which is intrinsically unstable, and the corresponding periodic state cannot last out, because any slight deviation from the point (2) will lead the oscillation to the stable state represented by either the point (1) or (3). It is also clear that one of the integral curves (i.e., the thick-line curve in Fig. 2) which contains the saddle point (2) separate the whole plane into two regions, in one of which all integral curves tend to the singularity (1), while in the other region to the singularity (3).

Figure 4 shows these regions in the r, θ plane where r is the amplitude and θ is the phase angle of v at an instant at which the external force is impressed, i.e.,

$$r=[x^2(0)+y^2(0)]^{\frac{1}{2}}, \quad \theta=\arctan[-y(0)/x(0)].$$

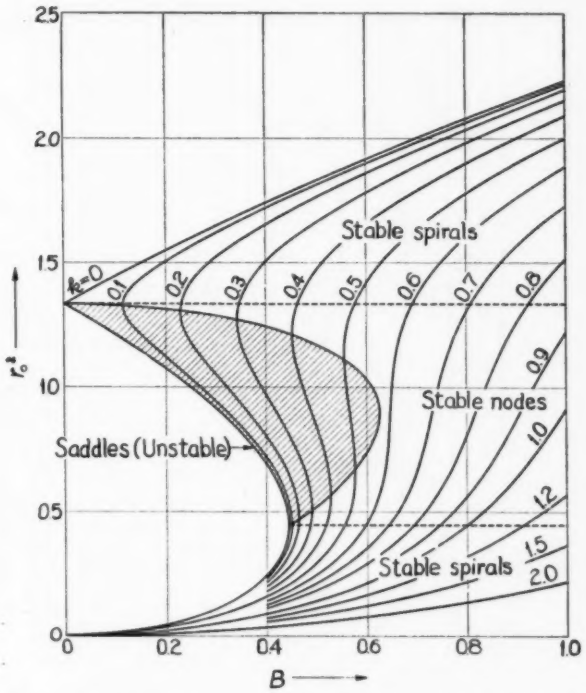


FIG. 1. Amplitude characteristics and the correlated singularities for the harmonic oscillations.

From this figure it is seen that an oscillation started with any initial conditions r and θ in the shaded region gets finally to the singularity (1), whereas an oscillation started from the remaining region gets finally to the singularity (3). It is also seen that the singular point (2) is situated on the boundary between these two regions.

In Fig. 3 we have only one stable node; in other words, we have a single periodic solution for any initial conditions prescribed.

It is obvious by Eqs. (2.3) that the amplitude and the phase angle of an oscillation in the neighborhood of the periodic state approach, with the lapse of time, that final state with damped sinusoids in the case when the corresponding singularity is a stable spiral, and approach to the final state with damped exponentials in the case when the singularity is a stable node. We can also see from Figs. 2 or 3 that the transient oscillation has higher frequency than that of the impressed force in the case when the initial value of r is comparatively large, because the point $x(\tau)$, $y(\tau)$ correlated with the said transient oscillation moves on an integral curve in the counter-clockwise sense around the origin 0.‡‡

4. Experimental Considerations

In this section we shall discuss the nonlinear oscillations governed by the original Eq. (1.1) and compare the

‡‡ Some further investigations into the geometry of integral curves (e.g., the limit cycle correlated with a quasi-periodic solution and Poincaré's "points singuliers de seconde espèce" correlated with periodic solutions under certain special cases) will be reported in the near future.

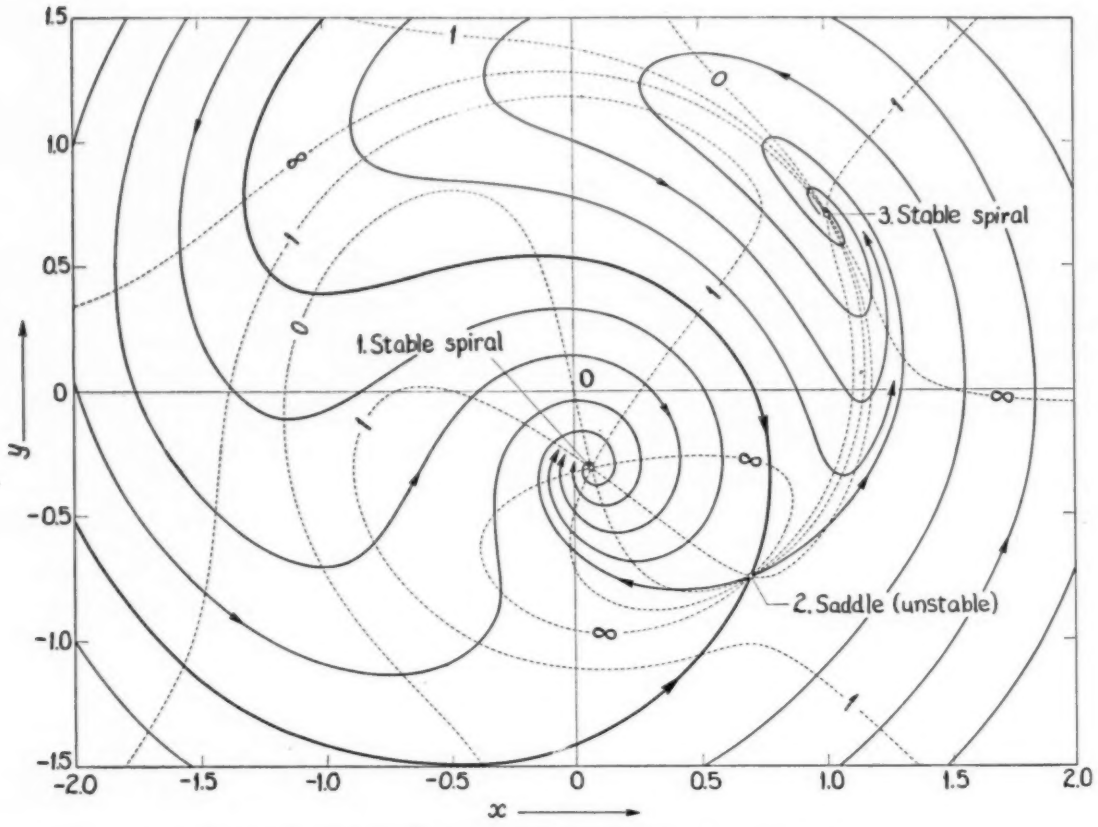


FIG. 2. Integral curves for the harmonic oscillations with three singularities.
Full lines: integral curves. Dotted lines: isoclines.

experimental results obtained with the theoretical results of the foregoing analysis.

The schematic diagram illustrated in Fig. 5 shows an electric circuit in which the nonlinear oscillations take place due to the saturable core inductance L under the alternating electromotive force $E\sin\omega t$. As shown in the figure, the resistance R is shunted to the capacitance C to render the circuit dissipative. With the notations of Fig. 5, we have

$$\left.\begin{aligned} n\frac{d\phi}{dt}+Ri_R&=E\sin\omega t,\\ Ri_R&=\frac{1}{C}\int i_C dt,\\ i&=i_R+i_C, \end{aligned}\right\} \tag{4.1}$$

where n is the number of turns of the coil and ϕ denotes the magnetic flux in the core.

Let us now introduce nondimensional variables u and v in place of i and ϕ by the following relations:

$$i=I_n\cdot u, \quad \phi=\Phi_n\cdot v, \tag{4.2}$$

where I_n and Φ_n are appropriate unit quantities of the current and the flux, respectively. Then, neglecting hysteresis, we may assume the saturation curve of the form

$$u=c_1v+c_3v^3+c_5v^5+\cdots, \tag{4.3}$$

where c_1, c_3, c_5, $\cdots$ are the constants characteristic to the core.§§

Although the units I_n and Φ_n can be chosen quite arbitrarily, it is preferable, for the brevity of calculation,

TABLE I. Singularities in Figs. 2 and 3.

Singular point	x_0	y_0	$\lambda_{1,2}$	Classification
Fig. 2, (1)	0.067	−0.310	$-0.100\pm0.423i$	stable spiral
Fig. 2, (2)	0.699	−0.748	0.170, −0.370	saddle (unstable)
Fig. 2, (3)	1.012	0.702	$-0.100\pm0.289i$	stable spiral
Fig. 3	0.983	−0.295	−0.082, −0.618	stable node

§§ For a usual transformer core, the coefficient c_1 of the linear term is very small compared to the other coefficients, but it may be enlarged by interposing an air gap in the core. The presence of c_1 causes no fundamental change in the experiment as well as in the theory. The presence of c_5, c_7, $\cdots$ exerts no serious influence so long as we deal with the harmonic oscillation. However, their presence is very significant in the case when we investigate the subharmonic oscillation. It is only mentioned here that the nonlinear relation given by Eq. (1.2) is most adequate to produce the subharmonic oscillation of order $\frac{1}{3}$, and such a characteristic may actually be obtained by connecting a number of inductance coils in series and adjusting the length of the air gap which is interposed in each core. The investigation into this problem requires a considerable space to discuss and so will be deferred in a later report.

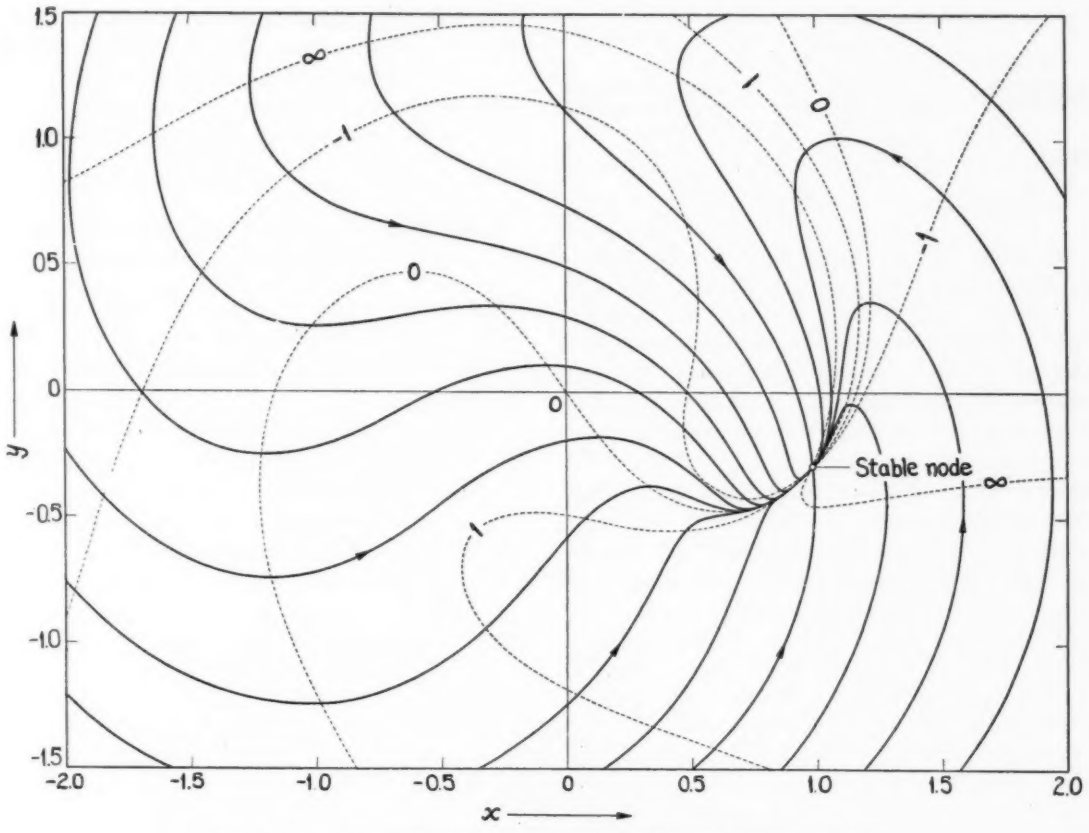


FIG. 3. Integral curves for the harmonic oscillation with one singularity.
Full lines: integral curves. Dotted lines: isoclines.

to fix them by the following relations, i.e.,

$$\left.\begin{aligned} n\omega^2 C\Phi_n &= I_n, \\ c_1+c_3+c_5+\cdots &= 1. \end{aligned}\right\} \tag{4.4}$$

Then, eliminating i_R and i_C from Eqs. (4.1) and making use of (4.2), (4.3), and (4.4), we obtain

$$\left.\begin{aligned} &\frac{d^2v}{d\tau^2}+k\frac{dv}{d\tau}+c_1v+c_3v^3+c_5v^5+\cdots=B\cos\tau, \\ &\text{where} \\ &\tau=\omega t-\arctan k, \quad k=\frac{1}{\omega CR}, \\ &B=(E/n\omega\Phi_n)(1+k^2)^{\frac{1}{2}}, \end{aligned}\right\} \tag{4.5}$$

which is of the same form as the original Eq. (1.1).

Now we proceed to describe the results of some experiments conducted by the author, as they are effective in verifying the foregoing analysis. As to the relation between the applied voltage E and the current i in Fig. 5, we have observed a good coincidence with the curves in Fig. 1, so that we have two stable periodic states in a certain range of the applied voltage. We shall distinguish these two states as resonant or nonresonant according as the amplitude of the oscillation is larger or smaller. It is therefore clear that the singularity (3) in Fig. 2 is correlated with the resonant state, while the singularity (1) is correlated with the nonresonant state.

In the foregoing analysis we have prescribed the

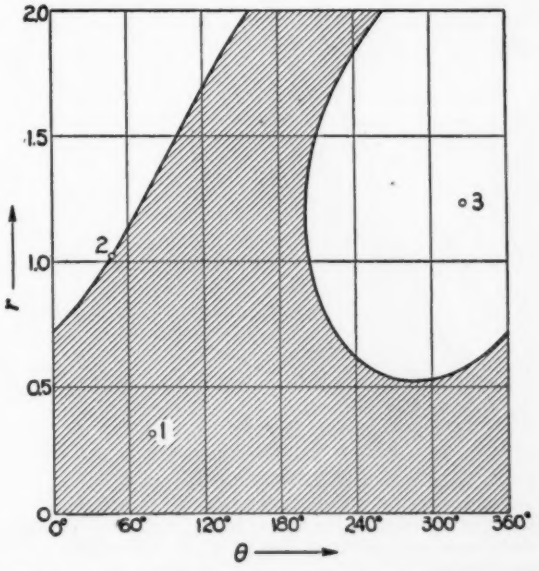


FIG. 4. Regions of initial conditions for the resonant oscillation (blank) and for the nonresonant oscillation (shaded).

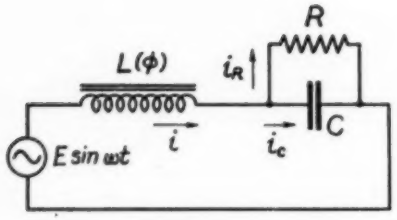


FIG. 5. Oscillatory circuit with nonlinear inductance.

initial conditions by the rectangular components of $v(0)$, i.e., the initial coordinates $x(0)$, $y(0)$ of a point in the x, y plane in Fig. 2. This is, however, not practical for our present experiment, so that we have prescribed the initial conditions instead by the initial charging voltage across the condenser and the phase angle at an instant at which the voltage had been impressed on the circuit.‖‖

Figure 6 shows the regions of initial conditions which lead either to the resonant or nonresonant state for several values of the applied voltage.¶¶ The shaded regions are for the nonresonant state, and as may be expected, these regions become constricted with increase of the applied voltage. These observed results are in accordance with the theoretical results of the preceding analysis (see Fig. 4) except for some slight discrepancies such as the branching of the shaded regions.

PART II. SUBHARMONIC OSCILLATIONS

5. Analysis of the Subharmonic Oscillations by Means of Integral Curves

Up to now we have considered the harmonic oscillations in which the frequency is the same as that of the external force. Permanent oscillations whose frequency is a fraction $\frac{1}{2}, \frac{1}{3}, \cdots, 1/n$ of that of the impressed force can, however, occur in nonlinear systems, in particular, in our case of the original Eq. (1.1). To this phenomenon the term "subharmonic oscillation" is usually applied.

As a typical example of the subharmonic oscillations, we shall treat only one special case, i.e., the subharmonic oscillation of order $\frac{1}{3}$ which is most frequently en-

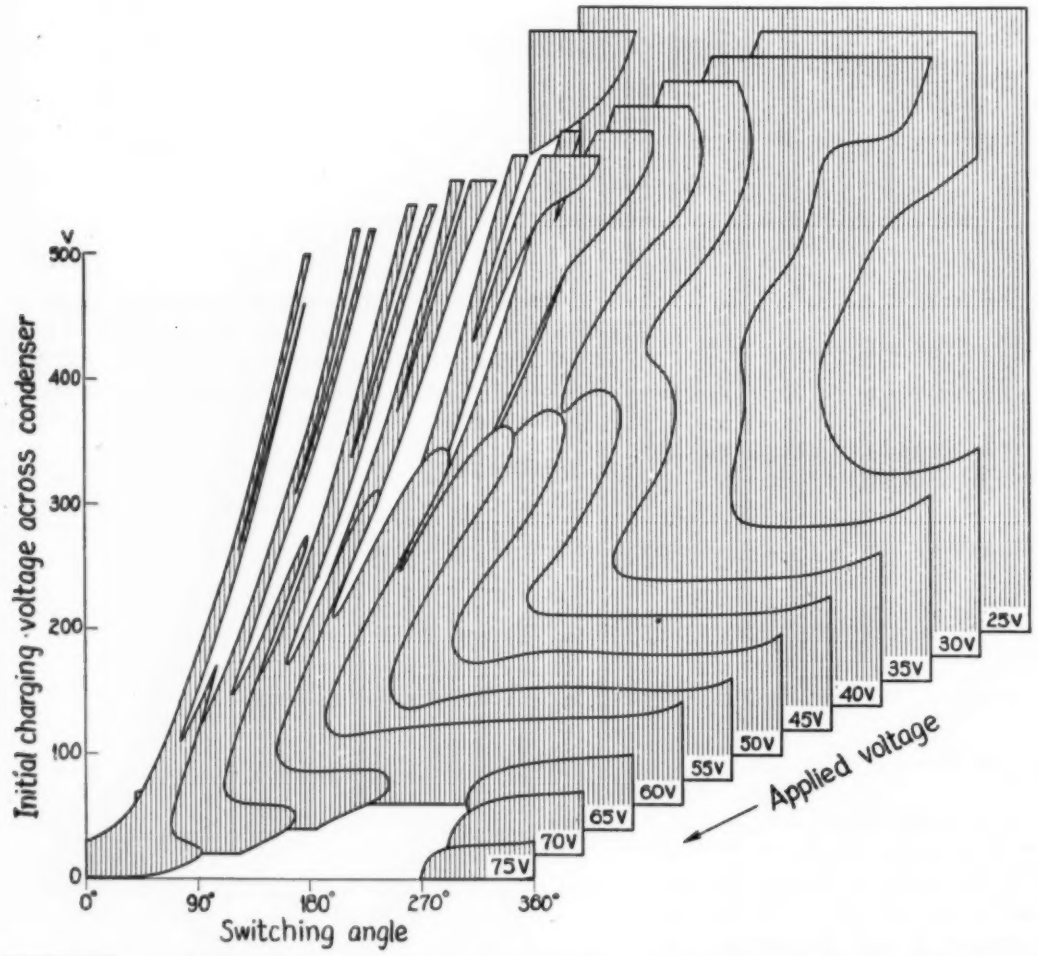


FIG. 6. Regions of initial conditions for the resonant oscillation (blank) and for the nonresonant oscillation (shaded) obtained under several values of the applied voltage.

‖‖ In order to switch in the circuit at any desired phase angle of the applied voltage, we have used an electronic switch consisting of two thyratrons connected in antiparallel. The grids of these thyratrons are further controlled electronically to effect the accurate timing.

¶¶ Before closing the circuit, the residual magnetism in the iron core must be demagnetized or at least kept at a certain constant value in each measurement.

countered in an oscillatory circuit containing saturable iron cores.***

We now consider the differential equation

$$\frac{d^2v}{d\tau^2}+k\frac{dv}{d\tau}+v^3=B\cos 3\tau, \tag{5.1}$$

in which it is to be noticed that $3(\omega)$ represents the angular velocity of the impressed force $B\cos 3\tau$.

For the subharmonic oscillation of order $\frac{1}{3}$, we may presume the solution of Eq. (5.1) in the form

$$v(\tau)=x(\tau)\sin\tau+y(\tau)\cos\tau+w\cos 3\tau. \tag{5.2}$$

According to Mandelstam and Papalexi,[5] the amplitude w of the component with the frequency of the external force is approximated by the following equation:

$$w=\frac{1}{1-3^2}B=-\tfrac{1}{8}B. \tag{5.3}$$

TABLE II. Singularities in Fig. 7.

Singular point	x_0	y_0	$\lambda_{1,2}$	Classification
Fig. 7, (1)	−1.123	−0.378	$-0.100\pm0.599i$	stable spiral
Fig. 7, (2)	0.889	−0.784	$-0.100\pm0.599i$	stable spiral
Fig. 7, (3)	0.234	1.162	$-0.100\pm0.599i$	stable spiral
Fig. 7, (4)	−0.858	0.209	0.365, −0.565	saddle (unstable)
Fig. 7, (5)	0.610	0.639	0.365, −0.565	saddle (unstable)
Fig. 7, (6)	0.248	−0.848	0.365, −0.565	saddle (unstable)
Fig. 7, (7)	0	0	$-0.100\pm0.500i$	stable spiral

Substituting (5.2) into (5.1) and separately equating the terms containing $\sin\tau$ and $\cos\tau$ to zero, we obtain

$$\left.\begin{aligned}\frac{dx}{d\tau}&=\tfrac{1}{2}[-kx+Ay+\tfrac{3}{4}w(x^2-y^2)]\equiv X(x,y),\\ \frac{dy}{d\tau}&=-\tfrac{1}{2}[Ax+ky+\tfrac{3}{4}w\cdot 2xy]\equiv Y(x,y),\end{aligned}\right\} \tag{5.4}$$

with

$$A=1-\tfrac{3}{4}r^2-\tfrac{3}{2}w^2,\quad r^2=x^2+y^2.$$

It should be remembered, however, that the same restrictions as those mentioned in Sec. 1 have to be assumed for the derivation of the above Eqs. (5.4).

We can now proceed in the same way as in the preceding case of the harmonic oscillations. Thus, the steady-state oscillations are correlated with the singularities determined by

$$X(x,y)=0,\quad Y(x,y)=0, \tag{5.5}$$

and the transient solutions can be discussed by the integral curves of the following equation:

$$\frac{dy}{dx}=\frac{Y(x,y)}{X(x,y)}, \tag{5.6}$$

where $X(x,y)$ and $Y(x,y)$ are given by Eqs. (5.4).

We now illustrate the integral curves of Eq. (5.6) for the following special case, i.e.,

$$k=0.2,$$
$$B=3.2\ [\text{accordingly } w=-0.4 \text{ by Eq. (5.3)}],$$

and find the regions of initial conditions under which the final periodic solutions get to the subharmonic response.

The singularities correlated with the periodic solutions are determined by Eqs. (5.5) and summed up in Table II. These singular points are shown in Fig. 7, together with the integral curves of Eq. (5.6) as drawn with the aid of isoclines. The three singular points (1), (2), and (3) which are correlated with the subharmonic oscillations are all equidistant from the origin $x=0$, $y=0$, and the angular distance between each pair of the singular points is equal to 120° or to one complete cycle of the impressed force. This is a plausible result for the subharmonic oscillation of order $\frac{1}{3}$.

An oscillation started with the initial conditions in the shaded regions gets ultimately to one of the singularities (1), (2), and (3) correlated with the subharmonic response, whereas any oscillation started from the remaining region gets ultimately to the singularity (7). Since $x=0$, $y=0$ at the singular point (7), no subharmonic oscillation takes place in this case.

Though they are not repeated here again, quite the same remarks as stated in Sec. 3 can also be applied to the subharmonic oscillations.

6. Experimental Investigations

Referring again to the electric circuit containing the saturable iron core (see Fig. 5), we shall compare the theoretical results of the foregoing analysis with experimental results.

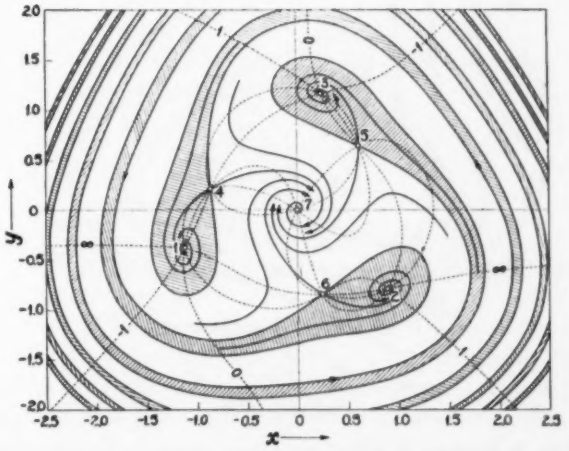


FIG. 7. Integral curves for the subharmonic oscillation of order $\frac{1}{3}$. Full lines: integral curves. Dotted lines: isoclines.

*** The subharmonic oscillations of other orders can be similarly treated provided that the nonlinear characteristics are appropriately assumed.

[5] L. Mandelstam and N. Papalexi, Z. Physik **73**, 227 (1932).

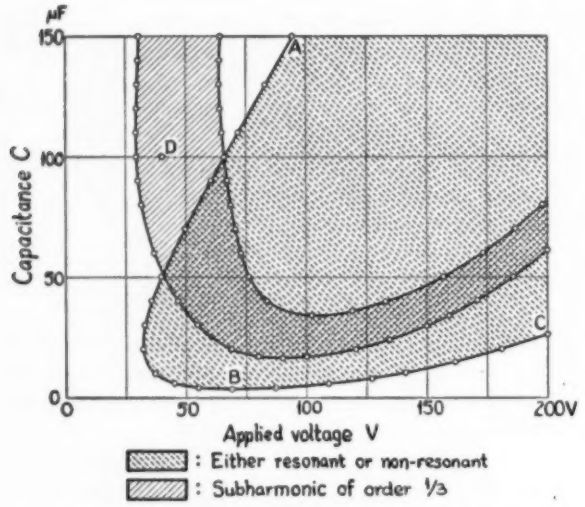


FIG. 8. Regions in which the oscillations of the different types sustain.

We have first determined the region in which subharmonic oscillations of order $\frac{1}{3}$ are maintained in stable states, by varying the applied voltage V and the capacitance C of the oscillatory circuit. The experimental result is shown in Fig. 8 where the said region is shaded by hatched lines. We have also determined the regions in which resonant and nonresonant harmonic oscillations may take place. The experimental result is shown with the boundary ABC in Fig. 8. An oscillation which is first in the nonresonant state with a comparatively low voltage transfers into the resonant state when the voltage is raised up to the boundary BC. On the contrary, a transition takes place from the resonant to nonresonant state when the voltage is reduced down to the boundary AB (see Fig. 1). Hence, in the interior of the boundary ABC shaded by dotted lines, either of the resonant and nonresonant harmonic oscillations may take place according to the initial conditions given.

As a result of this, we have two periodic states of oscillations, i.e., those of the nonresonant and subharmonic oscillations, under the circumstances represented by a point D, *viz.*,

$$C=100\ (\mu\text{F}), \quad V=40\ (\text{volt}).$$

Under these circumstances we have determined the regions of initial conditions which lead either to the nonresonant state or the subharmonic response. The experimental result is shown in Fig. 9, where the shaded regions are for the subharmonic response and are in good

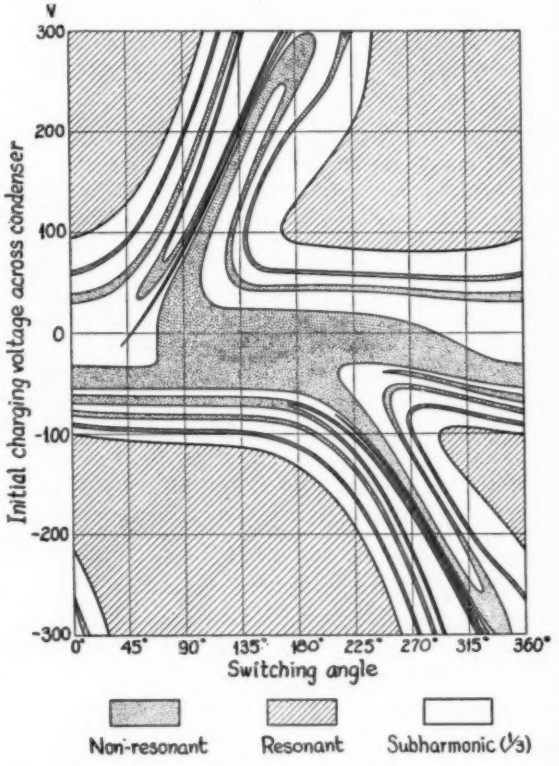


FIG. 10. Regions of initial conditions which lead to the different types of the periodic oscillations.

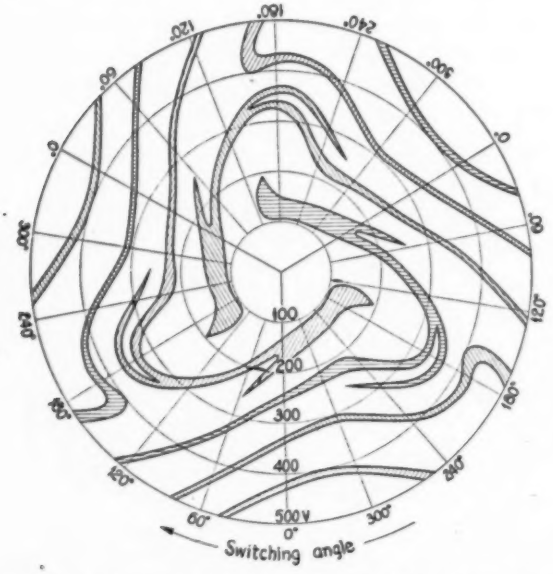


FIG. 9. Regions of initial conditions which lead to the subharmonic oscillation of order $\frac{1}{3}$. Radial scale: initial charging voltage across condenser.

accordance with the theoretical result given in the preceding section (see Fig. 7).

It is clear from Fig. 8 that the three kinds of periodic oscillations may sustain under certain circuit conditions.††† Figure 10 is an experimental result obtained under such a condition. Although the theoretical analysis has not yet been accomplished,‡‡‡ it may be ex-

††† These conditions are satisfied provided that the applied voltage and the capacitance are given within the region shaded by hatched lines and dots simultaneously.

‡‡‡ The geometrical method of analysis investigated in our paper fails for the case in which more than two states of periodic solutions may sustain in response to the different values of the initial conditions.

pected by analogy with Figs. 6 and 9 that the regions of subharmonic oscillations are constricted by the appearance of the regions of resonant harmonic oscillations and are distributed in narrower bands closely confined to one another.

ACKNOWLEDGMENTS

The present study was started under the guidance of Dr. R. Torikai, and the first preliminary report was published in 1943 in the Journal of the Institute of Electrical Engineers of Japan. The help and encouragement given the author by Dr. Torikai is gratefully acknowledged. The author's thanks are also due Dr. T. Matsumoto and Dr. S. Tomotika for their valuable criticisms and suggestions. In publishing the present paper the author is particularly grateful to Dr. Tomotika who kindly read the manuscript and gave unfailing advice of all kinds.

JOURNAL OF APPLIED PHYSICS VOLUME 24, NUMBER 2 FEBRUARY, 1953

Synthesis of Transfer Functions with Poles Restricted to the Negative Real Axis*

LOUIS WEINBERG
Hughes Research and Development Laboratories, Culver City, California
(Received September 15, 1952)

A procedure for the synthesis of general *RC* transfer functions by means of unbalanced networks is described. The transfer function need not be minimum phase but may have zeros anywhere in the complex plane except on the positive real axis. Use is made of the technique of zero shifting as in the Guillemin procedure; but the additional use of a network theorem divides the desired network into two parts, with a consequent reduction of the problem to two simpler problems. Zero shifting can now be performed *in two directions from within the total network*. The theorem plus a method of using fewer paralleled ladders yield a final network with fewer ladders and fewer elements than that given by the Guillemin procedure. In the illustrative example given, twenty-six elements are used, whereas the Guillemin procedure would use sixty-six.

INTRODUCTION

IN the past decade with the growth of applications of networks containing no inductance the *RC* synthesis problem has become increasingly important. This paper presents a solution to the problem by demonstrating a method for the realization of *RC* transfer admittances and voltage ratios as unbalanced networks. The given transfer function is not required to be minimum phase but may have zeros anywhere in the complex plane except on the positive real axis. Zero shifting and the paralleling of ladder networks are employed as in the Guillemin procedure (1); however, the additional use of a network theorem (2) and various other artifices simplifies the problem so that the final network has fewer ladders and fewer elements than the Guillemin network.

STATEMENT OF THE PROBLEM

We are given a rational function of the complex variable s,

$$\frac{p(s)}{q(s)}=\frac{s^n+a_{n-1}s^{n-1}+\cdots+a_1s+a_0}{s^m+b_{m-1}s^{m-1}+\cdots+b_1s+b_0}\ (m\geq n)$$
$$=\frac{(s-s_1)(s-s_3)\cdots(s-s_n)}{(s-s_2)(s-s_4)\cdots(s-s_m)}, \tag{1}$$

which we desire to realize within a constant multiplier H as the transfer voltage ratio†

$$K\equiv\frac{E_2}{E_1}=\frac{Hp(s)}{q(s)} \tag{2}$$

of the unbalanced *RC* network shown in Fig. 1.

For realization of a transfer function as an *RC* network, it is necessary and sufficient that the zeros of $q(s)$ all lie on the negative real axis. In order for the synthesis method of this paper to be applicable, the zeros of $p(s)$ may lie anywhere in the plane except on the positive real axis. Any given function that is minimum

---

* This paper was presented at the West Coast Institute of Radio Engineers Convention in August, 1952. It is based on a chapter in a thesis presented in partial fulfillment of the requirements for the Sc.D. degree at the Massachusetts Institute of Technology. The research was conducted under the supervision of Dr. E. A. Guillemin and was supported in part by the Air Materiel Command, the Army Signal Corps, and the U. S. Office of Naval Research.

[1] E. A. Guillemin, J. Math. and Phys. 22–42 (April, 1949).

[2] H. Adler, Electrical Engineering Master's thesis, Massachusetts Institute of Technology, 1948.

† The synthesis procedure for a voltage ratio is demonstrated; the transfer admittance case is essentially the same.

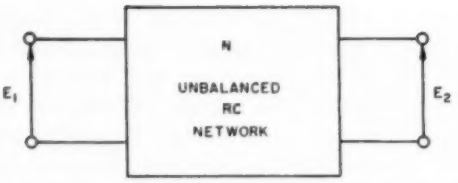


FIG. 1. Desired form of network for which $K=E_2/E_1$.

phase, i.e., has no zeros in the right half-plane, possesses a numerator with no negative coefficients. Though all the numerator polynomial coefficients may still be positive even for a function with right half-plane zeros, their positiveness is no longer guaranteed. However, as shown in Appendix I, a polynomial with some negative coefficients may be transformed to one with all positive coefficients by the use of surplus factors. In what follows it will therefore be assumed that the $p(s)$ has only positive coefficients, i.e., that the procedure for making all coefficients positive, if it is needed, has been carried out as a preliminary step in the synthesis.

The network to be synthesized may be divided into two parts, as shown in Fig. 2. For this network it is easy to show that (Appendix II)

$$K=y_{12a}K_b/(y_{22a}+Y_b), \tag{3}$$

where

$$\left.\begin{aligned} y_{12a}&=\frac{I_{sc}}{E_1}\\ K_b&=\frac{I_2Z}{E}=\frac{E_2}{E}\end{aligned}\right\}, \tag{4}$$

I_{sc} is the current obtained in a short circuit across terminals $a-b$, Y_b is the admittance seen looking to the right at $a-b$, and y_{22a} is the short-circuit driving-point impedance of network N_a.

The synthesis problem has thus been reduced to two simpler problems, for each of the partitioned networks may be worked on separately. To realize N_a, since it is characterized by *short-circuit* driving-point and transfer admittances, we may parallel ladder structures, each ladder being obtained by zero-shifting techniques which start from the right and work toward the left. (It is assumed that the reader is familiar with the Guillemin procedure so that the process of paralleling ladders and zero-shifting will not be explained here.) For N_b (i.e., N_b' and Z), on the contrary, we work from left to right and may use *only one ladder*. It is well known[1] that one ladder can realize any *RC* transfer function that has real and nonpositive zeros (including the origin and infinity); therefore only such zeros must be associated with the transfer function of N_b. The distribution of poles to the transfer functions of N_a and N_b determines the distribution of the partial fractions of the positive real denominator. It is necessary (and always possible) to make each ladder structure of N_a end on the left with a series branch while the ladder for N_b must end on the right with a shunt branch.

The synthesis will thus be accomplished if we first show how to modify the polynomials in Eq. (2) so that they can be identified with the driving-point and transfer characteristics of the two partitioned networks given by Eq. (3). This must be shown for both minimum phase and nonminimum phase functions. Then, after the identifications have been made, the networks N_a and N_b may be developed. Since the former is a group of paralleled ladders, it is of importance to present a method for obtaining a small number of ladders. In addition, it is shown that each of the component ladders may be associated with only a *part* of y_{22a} rather than the total of y_{22a}. The latter technique, when it is used, makes necessary the use of a new scheme for achieving a unity multiplier of y_{22a} for the paralleled network.

DETAILED ANALYSIS OF THE SYNTHESIS METHOD

The synthesis procedure for the nonminimum phase transfer function differs only slightly from that for the minimum phase. The latter will be considered in detail, and then we will indicate the slight change in point of view necessary for realizing the former.

Since the denominator of Eq. (2) must be identified with an *RC* driving-point admittance, we must divide it by a suitable polynomial q_1. This polynomial is chosen so that its zeros alternate with the zeros of q and the lowest magnitude zero occurs in q. The degree of q_1 is made one less than that of q. If the factored form of $p(s)$ shows that it has negative real zeros that occur within the intervals formed by the zero positions of q, then, for the purpose of reducing the number of elements in the final network, we choose zeros of q_1 at these points so that cancellations occur. Thus, as a simple illustration, if $q=(s+1)(s+3)(s+5)(s+7)(s+9)(s+11)$ and p is given by $(s+2)(s+4)(s^2+4s+9)$, then an acceptable q_1 is $(s+2)(s+4)(s+6)(s+8)(s+10)$.

When the numerator and denominator of Eq. (2) are divided by q_1, we obtain

$$K=(Hp/q_1)/(q/q_1), \tag{5}$$

whose new denominator q/q_1 may be identified as a positive real *RC* admittance.

Considering p/q_1 in its factored form, we make a distribution of the factors to the y_{12a} and K_b of Eq. (4). The distribution is made so that the transfer functions are realizable, a distribution which is always possible, and a small number of ladders and elements are required. Furthermore, as pointed out previously, K_b must be assigned only nonpositive real zeros. For example, if

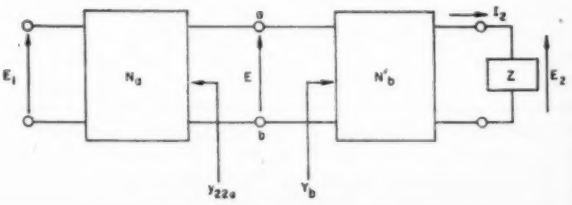


FIG. 2. Network after division into two parts.

$p=(s^2+2s+3)(s^2+3s+5)$ and q_1 is of the eighth degree, we would not choose

$$\left.\begin{aligned} y_{12a}&=\frac{(s^2+2s+3)(s^2+3s+5)}{\text{7th degree}}\\ Y_{12b}&=\frac{1}{\text{1st degree}}\end{aligned}\right\}, \tag{6}$$

for these identifications do not make use of the full power of the method. The synthesis would require three ladders with seventh-degree denominators as N_a and one ladder with a first-degree denominator for N_b. A better distribution would be

$$\left.\begin{aligned} y_{12a}&=\frac{(s^2+2s+3)(s^2+3s+5)}{\text{4th degree}}\\ Y_{12b}&=\frac{1}{\text{4th degree}}\end{aligned}\right\}, \tag{7}$$

for this requires only four ladders with fourth-degree denominators. Another good distribution is

$$\left.\begin{aligned} y_{12a}&=\frac{(s^2+2s+3)(s^2+3s+5)}{\text{3rd degree}}\\ Y_{12b}&=\frac{1}{\text{5th degree}}\end{aligned}\right\}. \tag{8}$$

We now need the q/q_1 broken into fractions, each of which is positive real. As is well known the partial fractions of q/q_1, since this function represents an *RC* admittance, will not have the required form. The partial fractions of q/sq_1, however, will have positive residues. Therefore we decompose q/sq_1 into partial fractions and then multiply both sides by s to obtain q/q_1 as the sum of positive real fractions.

We next identify as y_{22a} the sum of the fractions of q/q_1 that contain the poles of the previously identified y_{12a}, and perform a similar identification for Y_b and K_b. The constant term may be associated with either y_{22a} or Y_b. Similarly, if a pole at infinity does not occur in y_{12a} or K_b, then the first-degree term representing the pole at infinity may be identified as part of either y_{22a} or Y_b. The part of q/q_1 that contains poles not appearing in y_{12a} or K_b because of cancellations of equal factors of p and q_1 may be considered as part of Y_b.

In accordance with the preceding discussion, Eq. (5) has been transformed to

$$K=\frac{H(p_a/q_{1a})(p_b/q_{1b})}{(q_a/q_{1a})+(q_b/q_{1b})+(q_c/q_{1c})}, \tag{9}$$

where p_b/q_{1b} has only nonpositive real zeros and q_{1c} represents those factors of q_1 which have been canceled by equal factors of p. The Y_b term is identified with the sum of q_b/q_{1b} and q_c/q_{1c}, while y_{22a} is given by q_a/q_{1a}.

The ladder realizations may now proceed. In realizing Y_b as a single ladder we first remove the positive real q_c/q_{1c} as a shunt branch, so that its poles will not appear in K_b, and then zero shift so as to realize the remainder of Y_b and simultaneously insert the zeros of K_b. Network N_a may be realized by the Guillemin method except that, as demonstrated in Appendix III, a smaller number of ladders may be used. Thus, if n is the degree of p_a, the Guillemin procedure requires t ladders, where t is the smallest integer satisfying the relationship

$$t\geq(n+1)/2. \tag{10}$$

The new method requires a smaller t, namely, the smallest integer satisfying

$$t\geq(n-r+1)/2, \tag{11}$$

where r is the number of negative real roots present in the numerator p_a of y_{12a}.

As is also shown in Appendix III, as an alternative to some of the steps of the Guillemin method for realizing N_a, one can bring about further zero-pole cancellations in decomposing the numerator polynomial. This allows the association of only some of the positive real fractions of y_{22a} with each ladder, and hence decreases the number of required elements.

The above completes the consideration of the minimum phase function.

The nonminimum phase procedure is the same as the above except for an obvious slight change in point of view. In the factored form of p_a, the numerator of y_{12a}, we do not consider the quadratic factor containing complex roots in the right half-plane as a basic factor, but always regard it as associated with that factor necessary to render all its coefficients positive. The associated factor may, of course, be present in the given transfer function or may have been introduced as a surplus factor (Appendix I). Thus, for example, $p_a=(s^2-2s+7)(s+3)(s+1)$ would be considered in the form of two factors, $p_a=(s^3+s^2+s+21)(s+1)$.

ILLUSTRATIVE EXAMPLE

An *RC* network is desired to realize the transfer voltage ratio

$$K=\frac{E_2}{E_1}=\frac{(s^2+2s+2)(s+5.5)(s+10.5)}{(s+1)(s+2)(s+3)\cdots(s+10)}.$$

It is clear that the chosen example does not represent any desirable frequency characteristic. Because of the positions of the poles we can expect that a totally impractical spread of element values will be required, but this is of no concern to us at the present moment. The example has been chosen as a complicated one in order to illustrate as many of the synthesis details as possible and to show what is possible in reduction of the number of elements required by the Guillemin procedure.

We must identify the components in the equation

$$K=\frac{y_{12a}K_b}{y_{22a}+Y_b}.$$

To this end we perform the following manipulations on the given function:

$$K=\frac{\dfrac{(s^2+2s+2)(s+5.5)(s+10.5)}{(s+1.5)(s+2.5)\cdots(s+9.5)}}{\dfrac{(s+1)(s+2)\cdots(s+10)}{(s+1.5)(s+2.5)\cdots(s+9.5)}}$$

$$=\frac{\dfrac{(s^2+2s+.75+1.25)(s+10.5)}{(s+1.5)(s+9.5)}\dfrac{s+5.5}{(s+2.5)(s+3.5)\cdots(s+8.5)}}{D}$$

$$=\frac{\left[\dfrac{(s^2+2s+.75)(s+10.5)}{(s+1.5)(s+9.5)}+\dfrac{1.25(s+10.5)}{(s+1.5)(s+9.5)}\right]\dfrac{1}{(s+2.5)(s+3.5)(s+4.5)(s+6.5)(s+7.5)(s+8.5)}}{D}$$

$$=\frac{\left[\dfrac{(s+0.5)(s+10.5)}{s+9.5}+\dfrac{1.25(s+10.5)}{(s+1.5)(s+9.5)}\right]\dfrac{1}{(s+2.5)(s+3.5)(s+4.5)(s+6.5)(s+7.5)(s+8.5)}}{D},$$

where the denominator expanded in positive real fractions is

$$D=2.837732+s+\frac{0.556412s}{s+1.5}+\frac{0.087855s}{s+9.5}+\frac{0.471313s}{s+2.5}+\frac{0.392761s}{s+3.5}+\frac{0.328979s}{s+4.5}+\frac{0.275282s}{s+5.5}+\frac{0.227755s}{s+6.5}+\frac{0.183289s}{s+7.5}+\frac{0.138622s}{s+8.5}.$$

We now identify

$$y_{12a}=\frac{(s+0.5)(s+10.5)}{s+9.5}+\frac{1.25(s+10.5)}{(s+1.5)(s+9.5)}$$

$$=y_{12a}^{(1)}+y_{12a}^{(2)}$$

and

$$y_{22a}=\left\{s+\frac{0.047855s}{s+9.5}+c(2.837732)\right\}+\left\{\frac{0.556412s}{s+1.5}+\frac{0.0400s}{s+9.5}+(1-c)2.837732\right\},$$

$$=y_{22a}^{(1)}+y_{22a}^{(2)}$$

where c is a positive constant less than one. Its value will be chosen to make a zero occur in $y_{22a}^{(1)}$ at the desired position $s=-0.5$.

We further identify

$$K_b=\frac{1}{(s+2.5)(s+3.5)(s+4.5)(s+6.5)(s+7.5)(s+8.5)},$$

$$Y_b=\frac{0.275282s}{s+5.5}+\frac{0.471313s}{s+2.5}+\frac{0.392761s}{s+3.5}+\frac{0.328979s}{s+4.5}+\frac{0.227755s}{s+6.5}+\frac{0.183289s}{s+7.5}+\frac{0.138622s}{s+8.5}=\frac{0.275282s}{s+5.5}+\frac{e}{f},$$

where e/f represents all the poles other than the one at $s=-5.5$ combined into one quotient of polynomials.

The two ladders of network N_a are synthesized first.

We determine the constant c:

$$y_{22a}^{(1)}\Big|_{s=-0.5} = \left(s+2.837732c+\frac{0.047855s}{s+9.5}\right)\Big|_{s=-0.5} = 0$$

$$c = 0.177134$$

$$\therefore\ 2.837732c = 0.502659$$

$$(1-c)(2.837732) = 2.335073.$$

Next $y_{22a}^{(1)}$ is inverted and expanded in partial fractions

$$\frac{1}{y_{22a}^{(1)}} = \frac{0.994418}{s+0.5} + \frac{0.0055813}{s+9.550514}.$$

We now remove a series branch containing the total residue in the pole at $s=-0.5$. Then with the inverted remainder we create a zero at $s=-10.5$ by removing a capacitance equal to 16.2 farads. A final inversion yields

$$z = \frac{0.0055813}{0.909573(s+10.5)} = \frac{0.0061361}{s+10.5},$$

which when removed as the final series branch places a zero in $y_{12a}^{(1)}$ at $s=-10.5$.

For the synthesis of the second ladder of N_a, after calculating the value at $s=-10.5$ of $1/y_{22a}^{(2)}$ as 0.293754, we remove a resistance equal to this value:

$$z' = \frac{1}{y_{22a}^{(2)}} - 0.293754$$

$$= \frac{(s+1.5)(s+9.5)}{2.931485(s+1.210769)(s+9.374897)} - 0.293754$$

$$= \frac{0.133875(s+10.5)(s+3.068382)}{2.931485(s+1.210769)(s+9.374897)}.$$

Next from the inverted z' in the form

$$\frac{1}{z'} = 7.436910 + \frac{2.827242s}{s+10.5} + \frac{10.844648}{s+3.068382}$$

we remove a shunt branch equal to the total residue in the pole at $s=-10.5$. The inverted remainder yields

$$z'' = 0.054700 + \frac{1.820167}{18.281558s+22.819281},$$

from which the resistance equal to 0.0547 is removed. Again we invert the remainder and remove the total pole at infinity. Only a conductance remains which after inversion may be removed as the final series branch.

The transfer function K_b of the network N_b must have all its zeros placed at infinity and must possess no pole at $s=-5.5$. This is brought about by removing from Y_b the total residue in the pole at $s=-5.5$ as the initial shunt branch and then expanding the inverted remainder $z'''=e/f$ as the continued fraction

$$z''' = 0.573816 + \cfrac{1}{0.376494s + \cfrac{1}{3.048785 + \cfrac{1}{0.065548s + \cfrac{1}{30.81115 + \cfrac{1}{6.715667\times10^{-3}s + \cfrac{1}{2.322873\times10^{2} + \cfrac{1}{8.285600\times10^{-4}s + \cfrac{1}{4.325168\times10^{3} + \cfrac{1}{4.529744\times10^{-5}s + \cfrac{1}{4.892880\times10^{4} + \cfrac{1}{3.718739\times10^{-6}s}}}

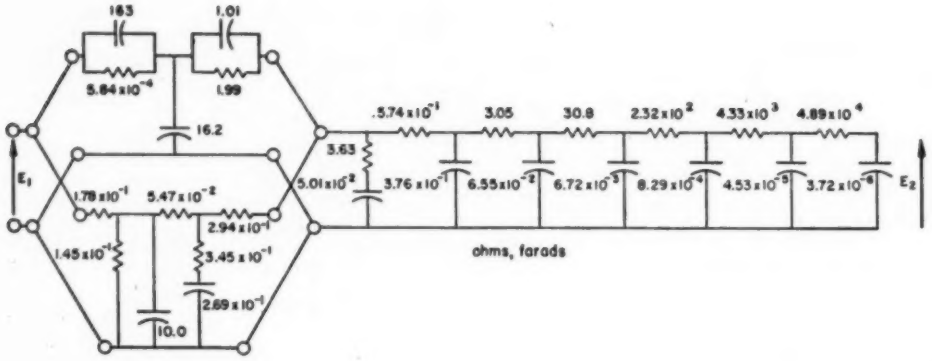


FIG. 4. Final network realized in illustrative example.

Allowing $s \rightarrow 0$ in the above equation and by inspection of the network we find that

$$\frac{A_1(0.5)(10.5)}{9.5} = \frac{1}{1.99},$$

$$A_1 = 0.909.$$

Similarly, A_2 is found to be 3.17.

Therefore

$$A_2/a_2 = 3.17/1.25 = 2.54,$$
$$A_1/a_1 = 0.909,$$

and we may write the achieved y_{12a} as

$$y_{12a} = \frac{A_1}{a_1}\left(\frac{a_1}{A_1}A_1 y_{12a}^{(1)} + \frac{a_1}{A_1}A_2 y_{12a}^{(2)}\right)$$

$$= 0.909(y_{12a}^{(1)} + 2.79 y_{12a}^{(2)}).$$

It is thus necessary to reduce the gain of the second ladder from 2.79 to 1.25, that is, multiply it by a factor $b = 0.448$. This is accomplished by breaking the final series resistance 7.98×10^{-2} of $y_{22a}^{(2)}$ into a final series R_1 followed by a shunt R_2, where

$$R_1 = \frac{7.98 \times 10^{-2}}{0.448} = 1.78 \times 10^{-1},$$

$$R_2 = \frac{7.98 \times 10^{-2}}{1 - 0.448} = 1.45 \times 10^{-1}.$$

The gain of the N_b network for K_b may also be determined by inspection. It is the reciprocal of the product of all the resistances and capacitances that realize the continued fraction expansion of z'''. This is calculated to be 1.63115×10^4. Another way of calculating this constant is by finding the constant term of the denominator polynomial of K_b. The fact that the first four figures obtained by each of these methods are equal constitutes a satisfactory check on the derived element values.

The final network, which is shown in Fig. 4, realizes the given transfer function within the constant multiplier $0.909 \times 1.63155 \times 10^4 = 1.48 \times 10^4$.

If we applied the Guillemin procedure to the same problem, we would require $(2n+2)$ elements for each of two ladders, $(2n+1)$ elements for the final ladder whose transfer function has a constant term as numerator, plus one resistance termination. Since the degree of the denominator polynomial n is 10, the grand total of required elements is 66. The network shown in Fig. 4 uses only 26 elements. There is thus a saving of 40 elements.

CONCLUSION

A method has been presented for realizing a nonminimum phase RC transfer function by means of an unbalanced network. The number of elements necessary for the realization is made small principally because of the use of four techniques: (a) the use of the partitioning theorem; (b) the use of a small number of paralleled ladders; (c) the use of zero-pole cancellation techniques; and (d) the possibility of association of each ladder in N_a with a driving-point function of lower degree than that of y_{22a}.

APPENDIX I. USE OF SURPLUS FACTORS

The given transfer function, if it is minimum phase, is composed of polynomials with only positive coefficients, since, as is well known, a necessary condition for the roots of a polynomial to lie in the left half-plane is that its coefficients be positive. A nonminimum phase function may still consist of polynomials with no negative coefficients, and thus allow the synthesis procedure of this paper to proceed directly; the positiveness of the coefficients, however, is no longer guaranteed. It is the purpose of this appendix to demonstrate the use of surplus factors for converting a transfer function with some negative coefficients into one with only positive coefficients.

Suppose we are given a transfer function to be synthesized,

$$K = p'(s)/q'(s), \tag{I.1}$$

where $p'(s)$ has some negative coefficients. We wish to convert this to an equivalent

$$K = \frac{p'(s)g(s)}{q'(s)g(s)} \equiv \frac{p(s)}{q(s)}, \tag{I.2}$$

where $p(s)$ now has only positive coefficients and $g(s)$ is an auxiliary polynomial or surplus factor with only negative real roots. The degree of this polynomial, as we shall see, is determined by the sector of the complex plane within which the roots of the original quadratic lie. The indicated multiplication by surplus factors does not change the *RC* character of the given function.

Since positive real zeros are excluded from the given transfer function, the negative coefficients of $p'(s)$ are caused by factors of the form $s^2-2\alpha s+\omega_0^2$, where $\alpha>0$, $\omega_0>0$, and $\omega_0^2>\alpha^2$. It is therefore sufficient to show how the negative sign may be removed from the quadratic and the type of procedure to be followed for determining a sufficient degree for the auxiliary polynomial.

Multiplication of the quadratic by a linear factor yields

$$(s^2-2\alpha s+\omega_0^2)(s+c)=s^3+(c-2\alpha)s^2+(\omega_0^2-2\alpha c)s+c\omega_0^2, \tag{I.3}$$

which, subject to the condition that all the coefficients shall be nonzero and positive, requires the satisfaction of the inequalities

$$\left.\begin{aligned} c&>2\alpha \\ \omega_0^2&>2\alpha c \end{aligned}\right\}. \tag{I.4}$$

The above can be combined to give

$$\frac{\omega_0^2}{2\alpha}>c>2\alpha \tag{I.5}$$

or

$$\omega_0^2>4\alpha^2, \tag{I.6}$$

if a real value of c is to exist. The condition (I.6) is equivalent to the requirement that the complex roots lie outside of the cross-hatched sector shown in Fig. 5, i.e., the sector formed by radial lines making an angle with the positive real axis of

$$\begin{aligned} \psi&=\cos^{-1}\frac{\alpha}{\omega_0}=\cos^{-1}\tfrac{1}{2} \\ &=60°. \end{aligned} \tag{I.7}$$

If we use a quadratic auxiliary polynomial we obtain the conditions to be satisfied

$$\left.\begin{aligned} c+d&>2\alpha \\ \omega_0^2-2\alpha(c+d)+cd&>0 \\ \omega_0^2(c+d)-2\alpha cd&>0 \\ \left(\frac{c+d}{2}\right)^2&\geq cd \end{aligned}\right\}. \tag{I.8}$$

Instead of solving these inequalities for the permissible sector in which the roots may lie, it is simpler and more instructive to show that c and d may be chosen to handle complex roots lying in a sector smaller than 60°. Suppose we choose

$$\left.\begin{aligned} c&=d \\ c+d&=3\alpha \end{aligned}\right\}. \tag{I.9}$$

Then the second inequality of (I.8) gives

$$\omega_0^2>1.5\alpha^2, \tag{I.10}$$

but this is not the limiting inequality, for the third inequality requires

$$\omega_0^2>3\alpha^2. \tag{I.11}$$

Thus, with the choice of c and d given by (I.9) a second-degree polynomial is adequate for roots lying outside of 54.7° sector.

It is heuristically evident (and it can be proved formally) that increasing the degree of the surplus factor will bring the permissible sector for roots closer and closer to the positive real axis.

Thus we see that the price we pay for synthesis of some nonminimum phase functions is an increase in the number of elements; other functions though nonminimum phase may, of course, have all positive coefficients without the use of surplus factors.

Finally it is clear why roots on the positive real axis are excluded, for by Descartes rule of signs a polynomial with a positive real zero must have at least one change of sign.

APPENDIX II. A NETWORK THEOREM

A theorem, which has been previously derived and called a partitioning theorem (2), but is basically an application of Thevenin's or Norton's theorem, is here explained with indications given of its application to synthesis problems.

Consider a network N shown in Fig. 6(a) where E_2 represents the open circuit voltage. After dividing this

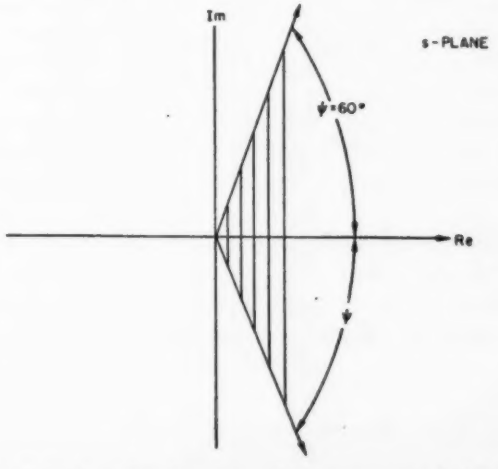


FIG. 5. Region of s plane from which roots are excluded if a linear auxiliary polynomial is used.

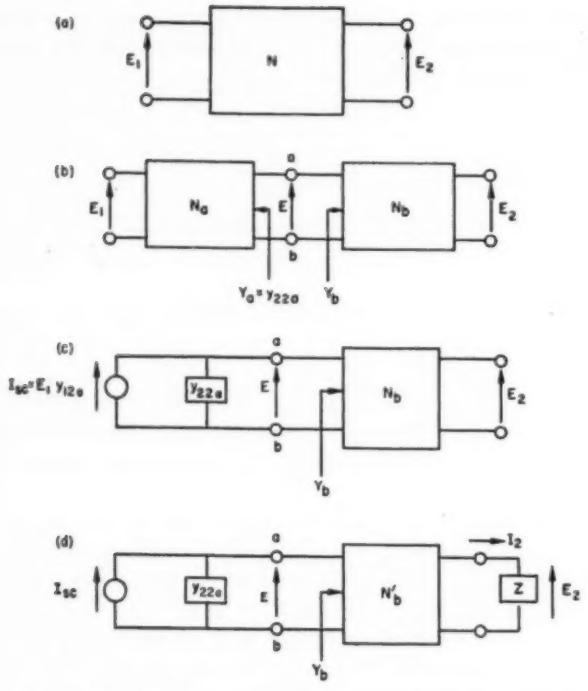


FIG. 6. Simplification of network N. (a) Original network. (b) Network decomposed into two parts. (c) Network after application of Norton's theorem to N_a. (d) Network with final branch shown explicitly.

network into any two smaller networks N_a and N_b, connected by a pair of wires, apply Norton's theorem to the network on the left, i.e., substitute an equivalent current source and shunt admittance. Finally, because of convenience for applications to synthesis, we show explicitly the final shunt arm of network N_b. This is indicated in Fig. 6(d), where N_b is the total right-hand network, comprising N_b' and Z, and E_2 is now the output voltage across the arbitrary termination Z. The complete sequence of steps is shown in Fig. 6.

With reference to the symbols defined in the figure, the following definitions are self-evident:

$$\left.\begin{aligned} Y_{12} &\equiv \frac{I_2}{E_1} \\ K &\equiv \frac{I_2 Z}{E_1} = \frac{E_2}{E_1} \\ Y_{12b} &\equiv \frac{I_2}{E} \\ K_b &\equiv \frac{I_2 Z}{E} = \frac{E_2}{E} = Z Y_{12b} \\ y_{12a} &\equiv \frac{I_{sc}}{E_1} \end{aligned}\right\}. \tag{II.1}$$

Also let Y_a and Y_b be the respective driving-point admittances of N_a and N_b for the directions indicated in Fig. 6(b). From its definition it is apparent that Y_a is equivalent to the short-circuit admittance y_{22a}. We now proceed to the derivation of the useful formulas.

From Fig. 6(d) we note that I_{sc}, the short-circuit current obtained by the Norton transformation, flows through Y_a plus Y_b to give rise to the voltage E; thus

$$\left.\begin{aligned} E &= \frac{I_{sc}}{Y_a + Y_b} \\ &= \frac{E_1 y_{12a}}{y_{22a} + Y_b} \end{aligned}\right\}. \tag{II.2}$$

Furthermore, since

$$I_2 = E Y_{12b}, \tag{II.3}$$

then substitution of Eq. (II.2) into Eq. (II.3) yields

$$I_2 = \frac{E_1 y_{12a} Y_{12b}}{y_{22a} + Y_b} \tag{II.4}$$

or

$$Y_{12} = \frac{I_2}{E_1} = \frac{y_{12a} Y_{12b}}{y_{22a} + Y_b}. \tag{II.5}$$

If we multiply the above by Z, we obtain

$$\left.\begin{aligned} K &= \frac{E_2}{E_1} = \frac{y_{12a} Y_{12b} Z}{y_{22a} + Y_b} \\ &= \frac{y_{12a} K_b}{y_{22a} + Y_b} \end{aligned}\right\}. \tag{II.6}$$

Analogous relationships can be derived for the transfer impedance Z_{12}.

It is evident that for an N_b network consisting of $R=1$ ohm, Eq. (II.5) reduces to the familiar

$$Y_{12} = y_{12}/(1+y_{22}), \tag{II.7}$$

which is the equation used, for example, in the Guillemin synthesis of RC networks. In the present paper the general equations are used to allow zero shifting *in both directions from within network* N. Thus it is clear that any advantages gained by use of these equations rather than Eq. (II.7) for RC synthesis derive from the fact that we have in effect used as a termination a network whose transfer and driving-point admittances are complicated functions of frequency rather than a termination of merely one mho. Similar remarks apply to RLC synthesis which will be treated in a later paper.

APPENDIX III. METHOD OF CHOOSING SMALL NUMBER OF PARALLELED LADDERS

Since N_a, one of the partitioned networks, consists of paralleled ladders, we here formulate a straightforward procedure for using a smaller number of ladders and fewer elements per ladder than are required by the

Guillemin procedure. It is recalled that in the latter method of *RC* synthesis (1), the given rational function

$$Y_{12}=\frac{p(s)}{q(s)} \tag{III.1}$$

is converted by means of the following steps:

$$Y_{12}=\frac{p}{q_1+q_2}=\frac{p/q_1}{1+(q_2/q_1)}, \tag{III.2}$$

so that identification may be made with the terms in

$$Y_{12}=\frac{y_{12}}{1+y_{22}}. \tag{III.3}$$

(We thus have the y_{12} and y_{22} which correspond to the y_{12a} and y_{22a} of N_a for the purposes of this discussion.) The numerator rational function $y_{12}=p/q_1$ is then divided into t components, each containing two successive terms of p as numerator and q_1 as denominator. Each component is identified with the $y_{12}^{(\mu)}$ of a network whose driving-point admittance is given by the *total* $y_{22}=q_2/q_1$. It is evident that the division of y_{12} in such a fashion requires that t ladders be used, where t is the smallest integer satisfying the relationship

$$t\geq(n+1)/2 \tag{III.4}$$

and n is the degree of p.

It is well known that only *one* ladder is needed to realize Y_{12} when the zeros of p are all real and nonpositive. However, though zeros of p are taken into account when *all* of them are negative real, the above procedure neglects to take similar advantage when only *some* of the zeros are negative real. The direct method explained below corrects this oversight.

If p is given in factored form, we note the number of negative real roots; if it is not given in factored form, we apply the Sturm test[3] to p to determine the number of negative real roots. Let this number be r. Then the number of ladders necessary, rather than being given by Eq. (III.4) is equal to the smallest integer satisfying

$$t\geq(n-r+1)/2. \tag{III.5}$$

The method is best explained by an example.

Suppose p is of the eighth degree and is given by

$$p=(s^2+2s+4)(s^2+3s+5)(s+6)(s+7)(s+8)(s+9)=(s^2+2s+4)(s^2+3s+5)B(s), \tag{III.6}$$

where $B(s)$ has been used for convenience. By the Guillemin procedure five ladders are required; by use of Eq. (III.5) t is given as only three. All we need do is divide the quartic containing the complex roots into three components and then multiply each of them by the four negative real roots. If the procedure of taking two terms at a time is used on the quartic, the breakdown is

$$p=(s^4+5s^3+15s^2+22s+20)B(s)=(s^4+5s^3)B(s)+(15s^2+22s)B(s)+20B(s). \tag{III.7}$$

The above completes the demonstration of how to determine the required number of ladders for N_a. It is not claimed that this method gives the minimum possible number of ladders; it is clear, however, that it gives a *straightforward* procedure for determining whether fewer ladders than that given by Eq. (III.4) can be used.

To reduce the number of elements in each ladder, the quartic factor is not broken down as in Eq. (III.7). When it is possible to make the roots of some of the components of p coincide with the chosen roots of q_1, we avail ourselves of this resource. The subsequent division by q_1 will cause a zero-pole cancellation in some of the $y_{12}^{(\mu)}$ components so that the number of partial fraction components of y_{22} that need be associated with these $y_{12}^{(\mu)}$ is smaller. As an example, suppose that for the p given by Eq. (III.6) the q_1 has zeros of interest at $s=-1$ and $s=-3$. We break down p so that the first component has a root at $s=-3$ and the second component one at $s=-1$. Thus instead of Eq. (III.7) we obtain

$$p=s^2(s^2+5s+6)B(s)+9(s^2+2.44s+1.44)B(s)+7.04B(s)=s^2(s+3)(s+2)B(s)+9(s+1)(s+1.44)B(s)+7.04B(s). \tag{III.8}$$

Thus, because of cancellations, the first ladder does not require the pole of y_{22} at $s=-3$ to be associated with it, while the second ladder can be realized without the pole at $s=-1$ in its driving-point function. If this association of some ladders with only a part of y_{22} is used, the Guillemin method of obtaining a unity multiplier for the y_{22} of the total network will not work, and it is then necessary to obtain the correct multiplier in a different manner which is discussed below. It is pointed out, however, that though the zero-pole cancellations discussed above are used, it is still possible to associate the total y_{22} with each ladder. In such a case the poles not present in the $y_{12}^{(\mu)}$ are removed as an initial shunt branch.

In associating with each component transfer function only those poles of y_{22} which the transfer function contains after possible cancellations, we do not associate the total fraction containing the pole but only part of it so that we can satisfy the relation

$$\sum_{\mu=1}^{t} y_{22}^{(\mu)}=y_{22}. \tag{III.9}$$

Thus, to take a simple example, for a network consisting

[3] L. E. Dickson, *New First Course in the Theory of Equations* (John Wiley and Sons, Inc., New York, 1939).

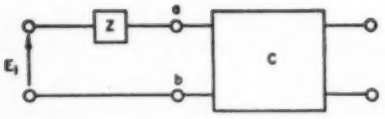


FIG. 7. Representation of a typical ladder of the network.

of three paralleled ladders, if

$$y_{22}=\frac{s}{s+2}+\frac{2s}{s+3}+\frac{3s}{s+4} \tag{III.10}$$

and if the pole at $s=-2$ occurs only in $y_{12}^{(2)}$, the pole at $s=-3$ occurs in all three $y_{12}^{(\mu)}$, and the $s=-4$ pole occurs in $y_{12}^{(1)}$, then acceptable driving-point functions for the respective ladders are

$$\left.\begin{aligned} y_{22}^{(1)}&=\frac{\frac{2}{3}s}{s+3}+\frac{3s}{s+4}\\ y_{22}^{(2)}&=\frac{s}{s+2}+\frac{\frac{2}{3}s}{s+3}\\ y_{22}^{(3)}&=\frac{\frac{2}{3}s}{s+3}\end{aligned}\right\}. \tag{III.11}$$

This artifice guarantees that the total driving-point admittance is the desired y_{22}, but we still have to guarantee that the constant multipliers of the $y_{12}^{(\mu)}$ are the desired ones.

Suppose that $a_\mu y_{12}^{(\mu)}$ is desired but $A_\mu y_{12}^{(\mu)}$ is achieved, so that

$$y_{12}=A_1y_{12}^{(1)}+A_2y_{12}^{(2)}+\cdots+A_ty_{12}^{(t)} \tag{III.12}$$

instead of the desired

$$y_{12}=a_1y_{12}^{(1)}+a_2y_{12}^{(2)}+\cdots+a_ty_{12}^{(t)}. \tag{III.13}$$

To correct this we compute each of the A_μ/a_μ to find the minimum ratio A_d/a_d. Factoring this ratio from the terms on the right side of Eq. (III.12),

$$y_{12}=\frac{A_d}{a_d}\left(\frac{A_1a_d}{A_d}y_{12}^{(1)}+\frac{A_2a_d}{A_d}y_{12}^{(2)}+\cdots\right. \tag{III.14}$$

$$\left.+a_dy_{12}^{(d)}+\cdots+\frac{A_ta_d}{A_d}y_{12}^{(t)}\right),$$

yields coefficients all except one of which are greater than the corresponding coefficients in Eq. (III.13), the dth coefficient being equal to the desired one. The

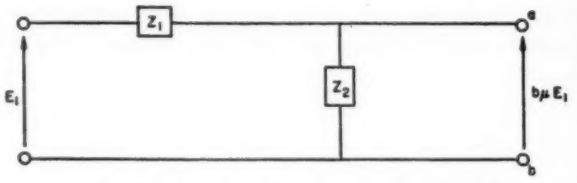


FIG. 8. Voltage division to obtain reduced gain.

factors which are necessary to reduce the coefficients are computed as

$$b_\mu=\frac{a_\mu}{A_\mu}\frac{A_d}{a_d}\quad\cdot\mu=1,2,\cdots t, \tag{III.15}$$

and the gain of the respective ladders must be reduced by this factor. This is easily done as shown by the following discussion.

Recalling that each ladder ends on the left in a series branch, suppose we represent the last branch by Z and the rest of the ladder by box C, as shown in Fig. 7. If we open circuit terminals $a-b$, we see looking to the left an impedance Z and a voltage E_1; we wish to see the same impedance Z but a reduced voltage $b_\mu E_1$. This can be achieved by the voltage division shown in Fig. 8, where

$$\left.\begin{aligned}\frac{Z_1Z_2}{Z_1+Z_2}&=Z\\ \frac{Z_2}{Z_1+Z_2}&=b_\mu\end{aligned}\right\}, \tag{III.16}$$

which yields in terms of the known b_μ and Z

$$\left.\begin{aligned}Z_1&=\frac{Z}{b_\mu}\\ Z_2&=\frac{Z}{1-b_\mu}\end{aligned}\right\}. \tag{III.17}$$

Most often the last branch Z is a resistance so that the procedure adds only one more element to each ladder.

Thus the final transfer function for the network is

$$y_{12}=\frac{A_d}{a_d}\left(a_1y_{12}^{(1)}+a_2y_{12}^{(2)}+\cdots+a_ty_{12}^{(t)}\right), \tag{III.18}$$

which is the desired one within a constant multiplier.

JOURNAL OF APPLIED PHYSICS VOLUME 24, NUMBER 2 FEBRUARY, 1953

Viscoelastic Properties as Functions of the Distribution of Activation Energies

W. JAMES LYONS
Chemical and Physical Research Laboratories, The Firestone Tire & Rubber Company, Akron, Ohio
(Received October 13, 1952)

1. Assuming that the elementary molecular deformation process conforms to the Maxwell model, and that the molecular elastic force G_i and viscous force η_i are functions (of unspecified forms) of the free energy of activation F^*, the following expressions for the dynamic modulus G_d and dynamic viscosity (internal friction) η_d are obtained:

$$G_d=\frac{1}{A}\int_0^\infty \frac{\omega^2\tau_i^2 G_i\phi(F^*)}{\omega^2\tau_i^2+1}dF^*,$$

and

$$\eta_d=\frac{1}{A}\int_0^\infty \frac{\eta_i\phi(F^*)}{\omega^2\tau_i^2+1}dF^*,$$

where A = area of sample, $\tau_i=G_i/\eta_i$, ω = vibration frequency, and $\phi(F^*)dF^*$ = the *number* of elementary processes having activation energies lying between F^* and F^*+dF^*.

2. By employing an expression relating the relaxation time τ_i with F^* for the elementary process, and adopting the so-called "box" distribution of relaxation times, the following explicit form for the distribution of activation energies is deduced:

$$\phi=\text{const}(1/kTF^*-1/F^{*2}),$$

where k = Boltzmann's constant and T = absolute temperature. When the box distribution, as represented by this explicit form for ϕ, is introduced into the foregoing expressions for G_d and η_d, the integrated results are found to predict temperature and frequency dependencies which are in gratifying agreement with experiment.

INTRODUCTION

THE distribution of relaxation times or rates (the so-called "relaxation spectrum") has been invoked in numerous researches to explain observed behavior involving the delayed responses of molecules or segments of molecules to imposed forces. It has been recognized also that the relaxation spectrum is but a reflection of a distribution of the more fundamental activation energies.[1–3] It is evidently desirable that such macroscopic properties as dynamic modulus and internal friction be expressed in terms of activation energies, and that the distribution of the latter, upon which also these properties depend, be known.

The objectives of this paper are (1) to derive relationships between the dynamic modulus and internal friction on the one hand, and a general expression for the activation energy distribution, on the other, and (2) to determine an explicit form for the distribution function, making use of an experimentally based expression in terms of relaxation times. While distributions of relaxation times, and their implications in mechanical behavior, have been exhaustively considered,[4] practically nothing specific has been written on the distribution of activation energies. It is well to note at this point that the term "distribution function," for relaxation times, has been used almost universally by writers in the high polymer and solid-state fields to designate the contribution to a modulus, compliance or stress made by all mechanisms having relaxation times in a particular infinitesimal range. In the terminology generally used in physics, a distribution function specifies the *number* of times values of a quantity lying within an infinitesimal range of values occur, or the *number* of existing elements having such values. In the present article the "distribution function" for activation energies is defined so as to conform to this preferred usage.

The pattern of the present development is to be as general as possible as regards the nature of the submicroscopic deformation process, to place the dependence of the process directly on activation energy, and to avoid the artificial concepts of springs and dashpots at the elementary level.

DEVELOPMENT OF GENERAL THEORY

Let us assume that for the elementary deformation process in a material, (1) the elastic and viscous components of the strain are directly additive, and (2) that the corresponding elastic (G_i) and viscous (η_i) forces are functions of a free energy of activation F^*. The elementary processes into which all macroscopic strains, whether tensile or shear, are ultimately resolvable are shear processes. For the elementary process, then, the rate of strain is

$$\frac{d\gamma}{dt}=\frac{1}{G_i(F^*)}\frac{df}{dt}+\frac{1}{\eta_i(F^*)}f, \qquad (1)$$

where γ is shear strain, f is imposed shearing *force*, $G_i(F^*)$ is the *spring constant* (shear) for the elementary

[1] W. Kauzmann, Revs. Modern Phys. 14, 12 (1942).
[2] M. L. Huggins, J. Appl. Phys. 21, 518 (1950).
[3] W. J. Lyons, J. Appl. Phys. 23, 287 (1952).
[4] For a comprehensive introduction to the field, see T. Alfrey, *Mechanical Behavior of High Polymers* (Interscience Publishers, Inc., New York, 1948), or B. Gross, J. Appl. Phys. 18, 212 (1947); 19, 257 (1948); 22, 1035 (1951).

process, and $\eta_i(F^*)$ is the corresponding viscous *force* (per unit rate of shear).

We now focus attention on a lamina of the material, bounded by planes transverse to the stress direction and having a depth (or thickness) equal, on the average, to the length of the elementary deformation unit. The aim here is to define a segment such that the only deformation forces acting across the transverse surfaces will be those arising in *single* elementary processes, and not in a number of such processes in series. The strains occurring in the lamina will be the same as those in the body as a whole. (The reader will readily recognize the lamina as the equivalent of a system of parallel Maxwell units.)

If now the lamina participates in steady harmonic vibration, each of the elementary processes occurring therein will follow the equation

$$\gamma=\gamma_{max}\exp i\omega t, \tag{2}$$

where γ_{max} is the shear strain-amplitude. Then

$$d\gamma/dt=i\omega\gamma_{max}\exp i\omega t=i\omega\gamma. \tag{3}$$

Similarly,

$$df/dt=i\omega f. \tag{4}$$

In general, γ and f will be out of phase with each other. Substituting in Eq. (1) from Eqs. (3) and (4),

$$i\omega\gamma=f\left[\frac{i\omega}{G_i(F^*)}+\frac{1}{\eta_i(F^*)}\right]. \tag{5}$$

Multiplying by $\eta_i(F^*)/G_i(F^*)\equiv\tau_i$ † and $-i\omega\tau_i(F^*)+1$,

$$i\omega\gamma\tau_i(-i\omega\tau_i+1)=\frac{f}{G_i}(i\omega\tau_i+1)\cdot(-i\omega\tau_i+1),$$

$$\omega^2\gamma\tau_i^2+i\omega\gamma\tau_i=\frac{f}{G_i}(\omega^2\tau_i^2+1). \tag{6}$$

Reinserting $d\gamma/dt$ for $i\omega\gamma$ [Eq. (3)] and rearranging terms, remembering that $G_i\tau_i=\eta_i$,

$$f=\frac{d\gamma}{dt}\frac{\eta_i}{\omega^2\tau_i^2+1}+\gamma\frac{\omega^2\tau_i^2G_i}{\omega^2\tau_i^2+1}. \tag{7}$$

It should be borne in mind that Eq. (7) remains an equation of motion for an elementary process.

Now the total force acting on the transverse area A of the lamina is

$$SA=\int_0^\infty f\phi(F^*)dF^*, \tag{8}$$

where S is shear stress, and $\phi(F^*)dF^*$ is the number of elementary processes having activation energies lying between F^* and F^*+dF^*. Thus $\phi(F^*)$ is, in the strict sense of the term, a distribution function for activation energies for a viscoelastic deformation process. Multiplying through by $\phi(F^*)dF^*$ in Eq. (7), we thus obtain

$$SA=\frac{d\gamma}{dt}\int_0^\infty\frac{\eta_i\phi(F^*)}{\omega^2\tau_i^2+1}dF^*+\gamma\int_0^\infty\frac{\omega^2\tau_i^2G\phi(F^*)}{\omega^2\tau_i^2+1}dF^*. \tag{9}$$

Since the strains in the lamina and the stresses on its transverse surfaces are the same as those in the macroscopic body, Eq. (9) can be taken as the equation of motion of viscoelastic matter in harmonic vibration.

The usual representation of a test material in dynamic experiments is a Voigt model, i.e., one having a single viscous and a single elastic element in parallel. The equation for this system is

$$S=(d\gamma/dt)\eta_d+\gamma G_d, \tag{10}$$

where S is the residual stress actually acting on the specimen after inertial forces (if any) associated with the apparatus have been subtracted from the applied force. The dynamic modulus G_d and the dynamic viscosity (or internal friction) η_d are defined by Eq. (10) and evaluated for the material as such a system by a corresponding dynamic experiment. If this experiment is conducted with harmonic vibrations, so as to impose on the elementary units similar vibrations, as indicated above, Eqs. (9) and (10) become identical. Hence we can conclude that

$$G_d=\frac{1}{A}\int_0^\infty\frac{\omega^2\tau_i^2G_i\phi(F^*)}{\omega^2\tau_i^2+1}dF^*, \tag{11}$$

and

$$\eta_d=\frac{1}{A}\int_0^\infty\frac{\eta_i\phi(F^*)}{\omega^2\tau_i^2+1}dF^*. \tag{12}$$

Consider again the thin lamina employed above. It will be apparent that the force exerted on a face of the lamina by an "instantaneous" elementary process will be wholly elastic. For a single process this force will be G_i, for unit shear. The instantaneous force over the whole face will be represented by

$$G_{inst}A=\int_0^\infty G_i\phi(F^*)dF^*, \tag{13}$$

where G_{inst} is the "instantaneous" modulus of the system.

It should be noted that Eqs. (11), (12), and (13), as derived by the foregoing development, are, within the limitations of the initial assumptions, of quite general applicability. Thus far, no explicit forms have been assumed or adopted for the functions $G_i(F^*)$, $\eta_i(F^*)$, their ratio τ_i, or $\phi(F^*)$. With an experimentally determined functional form for one or the other of the integrals in Eqs. (11)–(13), and assuming or being given the functions G_i and η_i, one is able, in principle at least, to derive the function ϕ. In principle also, the converse derivation of G_i and η_i is possible.

† The symbol $\tau_i(F^*)$ or τ_i is used here, for convenience, simply to represent the indicated ratio, and is not given physical significance in this development.

APPLICATION OF THEORY

1. Derivation of Distribution Function $\phi(F^*)$

An experimental expression for the integral in Eq. (13) is provided by Kuhn, Künzle, and Preissmann[5] and Tobolsky, Dunell, and Andrews,[6] in their analyses of mechanical property data on elastomers and other high polymers. While their results are in terms of relaxation times, an expression in activation energies can be derived, with recourse to the Eyring reaction-rate theory, As applied to shear, Tobolsky and co-workers' findings state that

$$G_{\text{inst}}=\int_0^\infty G_1(\tau)d\tau, \tag{14}$$

where $G_1(\tau)d\tau$ is the contribution of G_{inst} from Maxwell elements having relaxation times between τ and $\tau+d\tau$. They conclude from their analysis that

$$\left.\begin{aligned} G_1(\tau)&=\frac{G_0}{\tau}(\tau_l<\tau<\tau_m)\\ &=0(\tau<\tau_l,\ \tau>\tau_m)\end{aligned}\right\}, \tag{15}$$

where G_0 is a constant, and τ_i and τ_m are the lower and upper limits of a distribution of relaxation times. Hence

$$G_{\text{inst}}=\int_0^\infty G_0\frac{d\tau}{\tau}=\int_{\tau_l}^{\tau_m} G_0\frac{d\tau}{\tau}. \tag{16}$$

From Eqs. (13) and (16), reserving for later consideration the differences in the limits of integration, we get

$$G_i(F^*)\phi(F^*)dF^*=AG_0\frac{d\tau}{\tau},\quad(\tau_l<\tau<\tau_m). \tag{17}$$

As has been shown previously,[3] the reaction rate theory of viscoelastic properties, particularly as developed by Tobolsky, Powell, and Eyring,[7] leads to the expression

$$\tau_i=(h/2\pi^2F^*)\exp(F^*/kT), \tag{18}$$

where h is Planck's constant, k is Boltzmann's constant, T is absolute temperature, and the numerical value of F^* evidently refers to a single elementary process. Furthermore, from the latter authors' treatment it is deduced that

$$G_i(F^*)=\lambda_1 2\pi^2F^*/\lambda^2, \tag{19}$$

where λ is the distance between adjacent equilibrium positions, and λ_1 is the lateral distance between shearing molecular chains or segments.

[5] Kuhn, Künzle, and Preissmann, Helv. Chim. Acta **30**, 307 (1947).

[6] Tobolsky, Dunell, and Andrews, Textile Research J. **21**, 404 (1951).

[7] Tobolsky, Powell, and Eyring, edited by Burk and Grummitt, *Chemistry of Large Molecules* (Interscience Publishers, Inc., New York, 1943), p. 133.

Equation (18) leads to the relationship

$$d\tau_i/\tau_i=(1/kT-1/F^*)dF^*. \tag{20}$$

The quantities τ_i and τ differ in interpretation, the former being taken as a convenient symbol for a ratio, while τ is generally taken as a fundamental independent variable. Operationally, however, τ_i and τ are identical, both being the ratio of the viscous to the elastic constant in an elementary unit. Thus we may place $\tau=\tau_i$ and, from Eqs. (17), (19), and (20), get

$$\phi(F^*)dF^*=\frac{AG_0(1/kT-1/F^*)dF^*}{(\lambda_1 2\pi^2F^*)/\lambda^2},$$

or

$$\phi(F^*)=\text{const}(1/kTF^*-1/F^{*2}). \tag{21}$$

This function of F^* has a maximum at $F^*=2kT$. It is plotted in Fig. 1.

2. Range of Activation Energies

We wish now to consider what limits, if any, are to be imposed on F^*, corresponding to the limits τ_m and

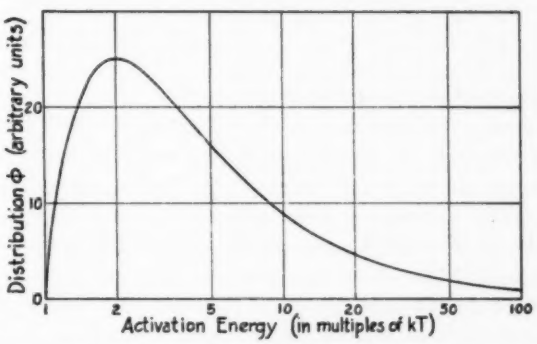


FIG. 1. Distribution function $\phi(F^*)$ for activation energies F^* for viscoelastic deformation, based on the "box" distribution of relaxation times τ_i and Eq. (18), relating τ_i and F^*.

τ_l, specified for Eq. (15). In Fig. 2 is plotted the graph of Eq. (18) for positive values of F^*. It will be seen that to each limit on τ_i there corresponds in general two limits on F^*. Two of these limits, F_a^* and F_b^*, however, lie in the range $(F^*<kT)$ in which ϕ, according to Eq. (21), is negative, and hence has no physical significance. The shape of the graph in Fig. 2 imposes a natural lower limit on τ_i, and this happens to fall at the point $F^*=kT$ at which (as just indicated) a lower limit on F^* is imposed by Eq. (21). While there is thus, according to Eq. (18), a lower limit on τ_i, corresponding to $F^*=kT$, there is no objective reason for taking this to be the τ_l deduced from mechanical experiments. It is true that the value of τ_i corresponding to $F^*=kT$ for ordinary temperature ($T=300°\text{K}$), being of the order of 10^{-14} sec, fulfils the condition that τ_m/τ_l be indeterminately large, deduced by Andrews, Hofman-Bang, and Tobolsky[8] for polyisobutylene. It is certainly

[8] Andrews, Hofman-Bang, and Tobolsky, J. Polymer Sci. **3**, 669 (1948). These writers state that $\tau_m/\tau_l=\infty(>10^4)$, but it is clear from the context of their article that merely a very large value for the ratio would meet their requirements.

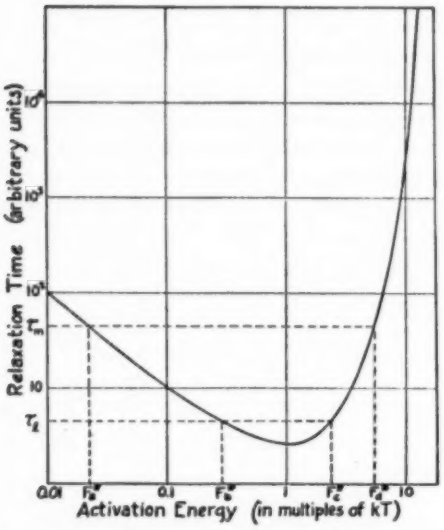


FIG. 2. Relaxation time τ as a function of activation energy F^*, according to Eq. (18). A logarithmic scale is used for F^*, to spread out the values of $F^*<kT$, for the sake of clarity.

to be expected, however, that the value of τ_l, as well as of τ_m, varies for the widely different high polymers, and probably for the different molecular species of a single polymer. Examination of the literature [8–12] tends to indicate that in general τ_m cannot be less than 10 sec. Corresponding to this value, $F_d^*=38.4kT$, according to Eq. (18). Expressing this value of F_d^* on a gram-mole basis at $T=300°K$, we obtain 23.0 kcal/gram-mole, which is of the order of (what are to be interpreted as) *average* activation energies, reported for viscoelastic deformation in phenolic resins and other polymers[13,14] and in polyamide monofils.[15] Considering, however, that at least one analysis[9] indicates relaxation times on the order of 10 sec to be intermediate, and that, furthermore many of the computed effective activation energies are greater than 23.0 kcal/gram-mole[13,14] it would appear that 10^5 or 10^6 sec is a more generally representative value for τ_m. The corresponding values of F_d^* are $47.9kT$ and $50.2kT$, by Eq. (18).* However, because of the indeterminacy in the values of τ_m, as well as τ_l, and the lack of general applicability, no corresponding limits on F^* are indicated in Fig. 1.

3. Dynamic Modulus Function

Referring to Eq. (11), we propose now to evaluate G_d in terms of F^*. From what has been said immediately above, it is evident that the limits of integration will be F_e^* and F_d^*. If the expressions for τ_i, G_i, and ϕ, given by Eqs. (18), (19), and (21), are substituted for these quantities in Eq. (11), a rather formidable integrand results. However, a readily manageable integrand is obtained by taking recourse to the use of $\tau=\tau_i$ as the variable, and making the substitution indicated by Eq. (17). The resultant expression is

$$G_d=G_0\int_{\tau_l}^{\tau_m}\frac{\omega^2\tau_i}{\omega^2\tau_i^2+1}d\tau_i. \tag{22}$$

The value of this integral is

$$G_d=\frac{G_0}{2}\ln\frac{\omega^2\tau_m^2+1}{\omega^2\tau_l^2+1}, \tag{23}$$

an expression obtained by Tobolsky, Dunnell, and Andrews.[6] Now, if the conditions of the dynamic experiment in which G_d is measured are such that

$$\tau_l\ll\frac{1}{\omega}\ll\tau_m, \tag{24}$$

then evidently

$$G_d\approx\frac{G_0}{2}\ln(\omega^2\tau_m^2/1).$$

We can now reintroduce F^*. Substituting for τ_m from Eq. (18), remembering that the corresponding limiting value of F^* is F_d^*, we have

$$G_d=\frac{G_0}{2}[\ln(\omega^2h^2/4\pi^4F_d^{*2})+F_d^*/kT]. \tag{25}$$

It is clear that for a given material (G_0 and F_d^* constant) at a single frequency ω, Eq. (25) expresses a hyperbolic relationship between G_d and T, the asymptotes being $G_d=(G_0/2)\ln(\omega^2h^2/4\pi^4F_d^{*2})$, and $T=0°K$. Such a relationship between dynamic modulus and temperature has been reported by Dillon, Prettyman, and Hall,[15] for natural and synthetic rubber stocks, and Hillier,[16] for polythene filaments. Ballou and Smith[17] report results on polyethylene terephtalate, and Nolle[11] has given results on various rubbers, which show a general decline in modulus with increasing temperature, though their results conform to Eq. (25) over only a part of the temperature range. Evidently, in these

[9] A. Tobolsky and H. Eyring, J. Chem. Phys. **11**, 125 (1943).
[10] T. Alfrey and P. Doty, J. Appl. Phys. **16**, 700 (1945).
[11] A. W. Nolle, J. Polymer Sci. **5**, 1 (1950).
[12] B. A. Dunell, thesis, "Dynamic Properties of Textile Fibers," Princeton University, 1949.
[13] N. A. de Bruyne, Proc. Phys. Soc. (London) **53**, 251 (1941).
[14] R. F. Tuckett, Trans. Faraday Soc. **40**, 449 (1944).
[15] Eyring, Alder, Rossmassler, and Christensen, Textile Research J. **22**, 223 (1952).

* In the initial paper on the application of reaction-rate theory to viscoelastic phenomena, H. Eyring (J. Chem. Phys. **4**, 283 (1936)) included in his equations a transmission coefficient κ, the probability that once having crossed a potential barrier, the deformation unit will not return to its initial state. The value of κ has always been assumed to be unity. This value was implicitly used in the development leading to Eq. (18). Evidently, from the definition of κ, it is restricted to the values $0\leqslant\kappa\leqslant1$. Introduced into Eq. (18) κ would appear in such a position that for values $\kappa<1$, larger values of τ_i would result, for a given F^*. Thus it is conceivable that the above estimates of relaxation times or corresponding activation energies are amendable upward or downward, respectively.

[15] Dillon, Prettyman, and Hall, J. Appl. Phys. **15**, 309 (1944).
[16] K. W. Hillier, Proc. Phys. Soc. (London) **B62**, 701 (1949).
[17] J. W. Ballou and J. C. Smith, J. Appl. Phys. **20**, 493 (1949).

cases, with decreasing temperature, crystallization or orientation effects enter, so that the structural parameters G_0 and F_d^* are altered.

It may be noted further that Eq. (25) is in good qualitative agreement with the increase in dynamic modulus with increasing logarithmic frequency, at constant temperature, found by Nolle.

In all these researches the macroscopic elastic and viscous *stresses* (rather than strains) are taken to be directly additive; i.e., the experimental material is treated implicitly as a single Voigt model. Thus the experimental moduli are computed for systems which are strictly represented by Eq. (10), and for which, therefore, Eq. (11) should hold.

4. Internal-Friction Function

By making in Eq. (12) the same substitutions as were made in Eq. (11) at the outset of the foregoing section, one obtains for the dynamic internal friction (viscosity),

$$\eta_d = G_0 \int_{\tau_l}^{\tau_m} \frac{d\tau_i}{\omega^2 \tau_i^2 + 1}. \tag{26}$$

The integrand here is of familiar form, yielding the evaluation

$$\eta_d = (G_0/\omega)(\tan^{-1}\omega\tau_m - \tan^{-1}\omega\tau_l). \tag{27}$$

For the conditions given by (24), $\tan^{-1}\omega\tau_l \approx 0$. In the series expansion of $\tan^{-1}\omega\tau_m$, the higher powers of $1/\omega\tau_m$ may be dropped, since, by the assumption of (24), $1/\omega\tau_m \ll 1$. Thus

$$\eta_d = (G_0/\omega)(\pi/2 - 1/\omega\tau_m + \cdots). \tag{28}$$

By virtue of Eq. (18),

$$\eta_d = (G_0/\omega)[\pi/2 - (2\pi^2 F_d^*/\omega h)\exp(-F_d^*/kT)]. \tag{29}$$

Equation (29) indicates that over a limited range of frequencies ω, at constant temperature, $\eta_d\omega$ = constant, an observation that has been made by a number of investigators.[15,18,19] On the other hand, according to Eq. (29), for materials having comparatively low upper limits on the activation energies, so that the second term in the bracket of Eq. (29) becomes influential, $\eta_d\omega$ would be expected to increase slowly with an increase in frequency, at constant temperature. This behavior has been found in Hevea rubber gum stocks, but is absent in the harder tread stocks.[15]

The relationship expressed by Eq. (29) which provides a unique test of the application of the present development is that between η_d and T. For temperatures sufficiently far removed from $T = 0°K$, at constant frequency, Eq. (29) predicts a decline of $\eta_d\omega$ with T, according to a hyperbola-like curve which is asymptotic to the value $\eta_d\omega = G_0[\pi/2 - 2\pi^2 F_d^*/\omega h]$. Behavior of this general type has been found by Nolle[11] and Dillon, Prettyman, and Hall[15] for natural and synthetic rubber vulcanizates.

DISCUSSION

The distribution $\phi(F^*)$ given by Eq. (21), as can be readily seen in Fig. 1, is not symmetrical. *A priori* one would expect the distribution to be symmetrical about some mean value. Kauzmann[1] cites three different distributions of relaxation rates (for dielectric polarization), all of which supposedly represent symmetrical distributions of activation energies. With the empirical adjustment of parameters in the distribution functions satisfactory agreement with experiment is claimed. In the present case the distribution represented by Eq. (21) is a direct translation of the "distribution" of τ defined by Eq. (15). This so-called distribution expressed as a function of log τ (the "box" distribution) is symmetrical, since the function has a constant value between the limits τ_l and τ_m. However, it is symmetrical in the contributions of the elements to the instanteous modulus G_{inst}, rather than in the *number* of elements having log τ values falling within consecutive infinitesimal intervals. An expression for τ which has been frequently postulated is the Arrhenius equation:

$$\tau = \tau_0 \exp(F^*/kT), \tag{30}$$

where τ_0 is a constant. Even if this were used instead of Eq. (18) in the foregoing development, the *number* distribution $\phi(F^*)$, as initially defined (Eq. (8)), would still not be symmetrical when applied to the box distribution.

It is worthy of reiteration that the general relationships given by Eqs. (11), (12), and (13) are free from objections to particular forms of $\phi(F^*)$ based on the interpretation of experimental data. Whether a symmetrical form for $\phi(F^*)$ can be found, such as to yield an approximation to the box distribution when expressed in appropriate terms, remains to be investigated.

ACKNOWLEDGMENT

To the research management of The Firestone Tire & Rubber Company the author is pleased to extend his appreciation for facilitating publication of this contribution.

[18] W. J. Lyons and I. B. Prettyman, J. Appl. Phys. **19**, 473 (1948).

[19] B. A. Dunell and J. H. Dillon, Textile Research J. **21**, 393 (1951).

Letters to the Editor

Shot Noise in Germanium Filaments

R. H. MATTSON AND A. VAN DER ZIEL
Department of Electrical Engineering, University of Minnesota, Minneapolis, Minnesota
(Received October 13, 1952)

IN a recent paper Herzog and van der Ziel[1] reported measurements on the shot noise of an n-type Ge-filament of 10 000 ohms. They found that between 1-1600 kc the noise could be represented by the formula:

$$n=1+AI^2/f+BI^2/[1+(f/f_0)^2], \quad (1)$$

where $f_0=150$ kc, $AI^2=9.3\times10^5$, and $BI^2=54$ for a current I of 3 mA. n is the noise ratio (=total noise power over thermal noise power); the second term in (1) is interpreted as excess noise the third one was interpreted as shot noise. A smaller term $CI^2/[1+(f/f_1)^2]$ also seemed to be present but the measurements above 500 kc were not sufficiently accurate and did not extend to sufficiently high frequencies to establish this with certainty.

The accuracy of the method of measurement has now been improved and the measurements have been extended to higher frequencies. Our results for the same filament indicate that formula (1) with the above values for f_0, AI^2 and BI^2 can account for all measurements and that the term $CI^2/[1+(f/f_1)^2]$ does not exist (Fig. 1).

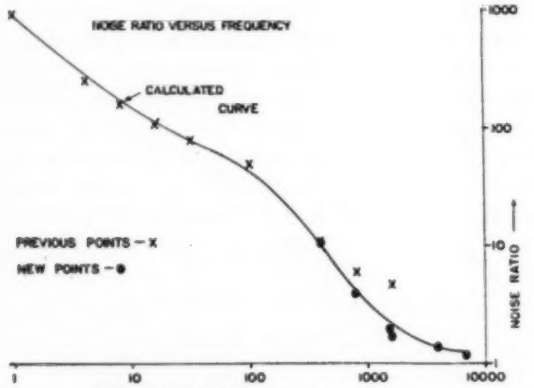


FIG. 1. Noise ration n as a function of frequency. Full drawn curve represents formula (1) with $f_0=150$ kc, $AI^2=9.3\times10^5$ and $BI^2=54$; Crosses represent previous data, circles represent new and more accurate data; the previous measurement at 1600 kc is now shown to be in error.

The section of the filament that was used for the experiments had a length d of 0.52 cm, a cross-sectional area of 8.4×10^{-4} cm², and a resistance of 10 000 ohms, corresponding to a resistivity of about 16 ohm cm which is relatively close to the intrinsic resistivity of about 60 ohm cm at room temperature.

Earlier measurements by Montgomery and Shockley[2] had shown that a major part of the excess noise (the term AI^2/f) at low frequencies was due to holes; Herzog and van der Ziel therefore assumed that the observed shot effect was due to holes too. In order to test this we compare the observed noise with theory.[3]

According to theory the noise ratio in a semiconductor in which the current is partly carried by electrons and partly by holes is

$$n_s=(e/kT)(V_0^2/d^2)\left[\frac{\alpha\mu_e\tau_e\sigma_e/(\sigma_e+\sigma_h)}{1+\omega^2\tau_e^2}+\frac{\beta\mu_h\tau_h\sigma_h/(\sigma_e+\sigma_h)}{1+\omega^2\tau_h^2}\right] \quad (2)$$

if the numbers of electrons and holes are fluctuating independently. μ_e and μ_h are the electron and hole mobilities, τ_e and τ_h are the electron and hole lifetimes, σ_e and σ_h are the electron and hole conductivities, respectively; V_0 is the dc voltage across the sample. If N_e is the number of electrons and N_h the number of holes in the sample, then α and β are defined by the relations

$$\overline{(N_e-\bar{N}_e)^2}=\alpha\bar{N}_e; \quad \overline{(N_h-\bar{N}_h)^2}=\beta\bar{N}_h. \quad (2a)$$

Our experiments apparently require that $\tau_h\simeq\tau_e$ in good approximation; the value $f_0=150$ kc corresponds to $\tau_e=1.1\times10^{-6}$ sec so that (2) may be written:

$$n_s=(e/kT)(V_0^2\mu_e\tau_e/d^2)\times[\alpha\sigma_e/(\sigma_e+\sigma_h)+\beta(\mu_h/\mu_e)\sigma_h/(\sigma_e+\sigma_h)]/(1+\omega^2\tau_e^2). \quad (3)$$

As $\mu_e=3600$ cm/sec per volt/cm and $\mu_h=1700$ cm/sec per volt/cm we find an intrinsic hole resistivity of 190 ohm cm. As our sample had a resistivity of 16 ohm cm, we find $\sigma_h/(\sigma_e+\sigma_h)=0.09$ and $\sigma_e/(\sigma_e+\sigma_h)=0.91$. For $I=3$ mA and hence $V_0=30$ volt we find $(e/kT)(V_0^2\mu_e\tau_e/d^2)=490$. As the measured value of n_s at low frequencies was 54, we obtain by substituting into (3)

$$0.91\alpha+0.04\beta=54/490=0.11. \quad (4)$$

Now if all electrons came from donor levels one would expect $\alpha\simeq0$ at room temperature, which would require $\beta=2.6$. This would mean that the holes were generated in bunches, which is not a very likely process. If all electrons would originate from deeper lying traps, one might expect a value of α of the right order of magnitude (see below); if an important part of the electrons came from pair generation, one would expect a much larger value for α than the maximum value of $=0.12$ which would occur for $\beta=0$.If all holes were generated by hole-electron pair creation one would expect $\beta=1$, otherwise $\beta<1$. The minimum value of α would occur for $\beta=1$, in which case $\alpha=0.08$.

The final conclusion is therefore that both carriers must have about the same life time of 1.1×10^{-6} sec and that the main part of the shot noise is due to electrons, though some part of the shot noise may be caused by the holes. In view of the short life time it seems possible that the electrons and holes come from (surface) traps. Recent experiments have shown that trapping centers occur at a depth of about 0.22 ev below the bottom of the conduction band.[4] This would give the right order of magnitude for α, for according to statistical mechanics[3]

$$[\bar{n}_1(1-\alpha)/\alpha]^{\frac{1}{2}}=3\times10^9\exp[-e\Delta E/(2kT)]. \quad (5)$$

Inserting $\Delta E=0.22$ ev and $\bar{n}_1$ (=the density of free electrons) $\simeq10^{14}$ (as follows from the conductivity) gives a value of α which has the right order of magnitude.

[1] G. B. Herzog and A. van der Ziel, Phys. Rev. **84**, 1249 (1951).
[2] H. C. Montgomery and W. Shockley, Phys. Rev. **78**, 646 (1950).
[3] A. van der Ziel, J. Appl. Phys. **24**, 222 (1953).
[4] R. N. Hall, Phys. Rev. **87**, 387 (1952).

Shot Noise in Semiconductors

A. VAN DER ZIEL
Department of Electrical Engineering, University of Minnesota, Minneapolis, Minnesota
(Received October 8, 1952)

CONSIDER a block of semiconducting material of length d having two plane parallel electrodes at the ends. Let I_0 be the dc current flowing through the conductor and V_0 the dc voltage applied to the electrodes.

Let N_e be the number of free electrons and N_h the number of free holes in the sample and let N_e and N_h fluctuate such that

$$\overline{(N_e-\bar{N}_e)^2}=\alpha\bar{N}_e; \quad \overline{(N_h-\bar{N}_h)^2}=\beta\bar{N}_h; \quad (1)$$

the magnitude of α and β depends upon the conditions of the semiconductor. Let τ_e be the average lifetime of a free electron and τ_h the average lifetime of a free hole. Due to the electric field strength $E=V_0/d$ the electrons obtain a drift velocity μ_eV_0/d and the holes obtain a drift velocity μ_hV_0/d, where μ_e and μ_h are the electron and hole mobilities, respectively. We may then calculate the shot noise generated in the semiconductor by the flow of dc current.

According to Ramo's theorem a single carrier of charge e moving with a velocity u gives a current pulse $i(t)=eu/d$ in the outer circuit. Consequently, we have the dc current I_0

$$I_0=e\mu_e\bar{N}_eV_0/d^2+e\mu_h\bar{N}_hV_0/d^2=(eV_0/d^2)(\mu_e\bar{N}_e+\mu_h\bar{N}_h), \quad (2)$$

and the fluctuations in N_e and N_h produce a fluctuating current

$$I(t)=(eV_0/d^2)[\mu_e(N_e-\bar{N}_e)+\mu_h(N_h-\bar{N}_h)]. \quad (3)$$

The autocorrelation function of $I(t)$ is, if $(N_e-\bar{N}_e)$ and $(N_h-\bar{N}_h)$ fluctuate *independently:*

$$\overline{I(t)I(t+s)}=(eV_0/d^2)^2$$
$$\times[\mu_e^2\overline{(N_e-\bar{N}_e)_t(N_e-\bar{N}_e)_{t+s}}+\mu_h^2\overline{(N_h-\bar{N}_h)_t(N_h-\bar{N}_h)_{t+s}}]$$
$$=(eV_0/d^2)^2[\mu_e^2\alpha\bar{N}_e e^{-s/\tau_e}+\mu_h^2\beta\bar{N}_h e^{-s/\tau_h}], \quad (4)$$

if the length over which the electrons and holes drift during their life time is small in comparison with d.

The noise may be represented by a noise current generator $(\langle i^2\rangle_{Av})^{\frac{1}{2}}=[w(f)df]^{\frac{1}{2}}$ of infinite impedance in parallel to the conductor; the noise intensity follows from the Wiener-Khintchine theorem:

$$w(f)=4\int_0^\infty \overline{I(t)I(t+s)}\cos\omega s\,ds$$
$$=4(eV_0/d^2)^2[\alpha\bar{N}_e\mu_e^2\tau_e/(1+\omega^2\tau_e^2)+\beta\bar{N}_h\mu_h^2\tau_h/(1+\omega^2\tau_h^2)] \quad (5)$$

and the noise ratio n_s of the shot noise (= total shot noise power over thermal noise power) becomes

$$n_s=w(f)/(4kT/R)=(e/kT)(V_0^2/d^2)$$
$$\times\left[\frac{\alpha\bar{N}_e\mu_e^2\tau_e/(1+\omega^2\tau_e^2)+\beta\bar{N}_h\mu_h^2\tau_h/(1+\omega^2\tau_h^2)}{\mu_e\bar{N}_e+\mu_h\bar{N}_h}\right]$$
$$=(e/kT)(V_0^2/d^2)\left[\frac{\alpha\mu_e\tau_e\sigma_e/(\sigma_e+\sigma_h)}{1+\omega^2\tau_e^2}+\frac{\beta\mu_h\tau_h\sigma_h/(\sigma_e+\sigma_h)}{1+\omega^2\tau_h^2}\right], \quad (6)$$

where $\sigma_h=e\bar{n}_h\mu_h$ and $\sigma_e=e\bar{n}_e\mu_e$ are the hole and electron conductivity and $\bar{n}_h$ and $\bar{n}_e$ are the average hole and electron densities, respectively.

We apply this to various cases.

(a) *Electron conductors*, that is $\sigma_h=0$. Let N_0 represent the number of donor levels in the sample and $\alpha=(1-\bar{N}_e/N_0)$ the probability that the donor level is neutral, then according to probability theory

$$\langle(N_e-\bar{N}_e)^2\rangle_{Av}=N_0\alpha(1-\alpha)=\bar{N}_e\alpha, \quad \text{as} \quad \bar{N}_e=N_0(1-\alpha), \quad (7)$$

and consequently,

$$n_s=(e/kT)(V_0^2\mu_e\tau_e\alpha/d^2)/(1+\omega^2\tau_e^2). \quad (8)$$

This is small if τ_e is small or if α is close to zero. For $\alpha\simeq 1$ Eq. (8) reduces to a formula published previously.[1-3] A measurement of n_s as a function of frequency thus gives τ_e and α if μ_e is known, or τ_e and μ_e if α is known.[4]

According to statistical mechanics we have

$$\bar{n}_e/[n_0-\bar{n}_e]^{\frac{1}{2}}=[\bar{n}_e(1-\alpha)/\alpha]^{\frac{1}{2}}=3\times10^9e^{-e\Delta E/(2kT)}\ \text{cm}^{-\frac{1}{2}}, \quad (9)$$

where n_0 and $\bar{n}_e$ are the densities of donors and free electrons, respectively. Knowing $\bar{n}_e$ from the resistivity and α, one may calculate ΔE, the energy gap in electron volts between the donor levels and the bottom of the conduction band. If ΔE is very small, the shot effect will be negligible at room temperature (as nearly all donor atoms are ionized, the number of free electrons hardly fluctuates); it then only becomes important at low temperatures.

(b) *Mixed conductors*, such that only the number of holes can fluctuate at random ($\beta=1$) but the number of electrons does not fluctuate ($\alpha=0$),

$$n_s=(e/kT)(V_0\mu_h\tau_h/d^2)[\sigma_h/(\sigma_e+\sigma_h)]/(1+\omega^2\tau_h^2). \quad (10)$$

(c) *Intrinsic semiconductors* in which electrons and holes appear and disappear by surface trapping. In that case $(N_e-\bar{N}_e)$ and $(N_h-\bar{N}_h)$ will fluctuate independently, and the full formula (6) holds.

(d) *Intrinsic semiconductors*, in which electrons and holes disappear only by recombination. In that case the hole-electron pairs are generated and pairs disappear by recombination, so that $\bar{N}_e=\bar{N}_h$, $\tau_e=\tau_h$ and $(N_e-\bar{N}_e)=(N_h-\bar{N}_h)$; that is, the fluctuations are completely correlated. $\langle I(t)I(t+s)\rangle_{Av}$ then has to be recalculated from (3). If the hole-electron pairs are generated at random we have $\langle(N_e-\bar{N}_e)^2\rangle_{Av}=\bar{N}_e$ and a simple calculation shows that

$$n_s=(e/kT)[V_0^2(\mu_e+\mu_h)\tau_e/d^2]/(1+\omega^2\tau_e^2). \quad (11)$$

All above examples hold under the assumption that the average drift length of the electrons and holes is small in comparison with d. Davydov and Gurevich[3] have solved the case in which this assumption is dropped, but only for $\alpha=1$, $\beta=0$. The general case of arbitrary values of α and β has not yet been solved.

[1] J. Bernamont, Compt. rend. **198**, 1755, 2144 (1934); Ann. phys. **7**, 71 (1937).
[2] A. van der Ziel, Physica **16**, 359 (1950).
[3] B. Davydov and B. Gurevich, J. Phys. U.S.S.R. **7**, 138 (1943).
[4] R. L. Petritz, Phys. Rev. **86**, 660 (1950).

Noise in Gas Discharges

A. VAN DER ZIEL
Department of Electrical Engineering, University of Minnesota, Minneapolis, Minnesota
(Received October 8, 1952)

PARZEN and Goldstein[1] have shown that the available noise power in a gas discharge for a band width df is

$$P_a=kT_e df+(P_0\tau_0/\bar{N}_1)[1+2/(1+\omega^2\tau_0^2)]df, \quad (1)$$

where T_e is the electron temperature of the discharge, $\bar{N}_1$ the average number of electrons in the discharge, P_0 the dissipated dc power, and τ_0 the average time between two collisions.

The above formula was derived under the assumption that the individual current pulses of the electrons can be treated as independent. This is perhaps justified as far as thermal noise is concerned, for it is usually assumed that no correlation exists between the velocity of an electron before and after a collision (though this is in doubt[2]). But without question the assumption is invalid for shot noise, as this effect is due to the drift of the electrons in the field. Even after collisions the *same* electrons will perform the *same* drift; for any particular electron this will go on until it is taken out of circulation by being captured by an ion. The current pulses caused by the drift of the *same* electron should therefore not be treated as independent. This means that the above formula can only be correct for frequencies such that $\omega\tau_0\gg1$. For the probability of a free time between collisions in the interval between τ and $(\tau+d\tau)$ is

$$dP=\exp-(\tau/\tau_0)d(\tau/\tau_0), \quad (2)$$

so that the randomness of the free time τ will ensure that the subsequent current pulses of the same electron can be treated as independent if $\omega\tau_0\gg1$.

In order to see how this modifies the shot noise term (=the second term in (1)) we may apply the theory of shot noise in semiconductors, as the problems of shot noise in semiconductors and of the low frequency shot noise in gas discharges are identical. Let the gas discharge occur between plane parallel electrodes at a distance d. Let τ_1 be the average lifetime (=time between its release by a neutral atom and its recapture by an ion) of an electron and let during its lifetime the electron drift over a distance which is small in comparison with d. Let μ_e (ratio of the drift velocity and the field strength) be the electron mobility (cm/sec per volt/cm), V_0 the dc potential between the electrodes, $P_0=V_0^2/R$ the dissipated dc power, and

$$1/R=e\mu_e\bar{N}_1/d^2 \quad (3)$$

the dc conductivity of the discharge. If the probability that the atoms in the discharge are neutral is close to unity, as should be

expected, the contribution n_s of the shot effect to the noise ratio (= total noise power over thermal noise power) of the discharge is[2]

$$n_s=(e/kT)(V_0^2\mu_e\tau_1/d^2)/(1+\omega^2\tau_1^2)$$
$$=(1/kT)(P_0\tau_1/\bar{N}_1)/(1+\omega^2\tau_1^2). \quad (4)$$

The contribution to the available noise power P_a is $n_s \cdot kTdf$, and hence,

$$P_a=kT_0df+(P_0\tau_1/\bar{N}_1)df/(1+\omega^2\tau_1^2) \quad (5)$$

for $\omega\tau_0\ll 1$. The low frequency shot noise is therefore much larger than the high frequency shot noise. Equation (5) shows that the measurement of shot noise of a gas discharge may be used to determine the life time τ_1 of the electrons.

Equations (1) and (5) describe the cases $\omega\tau_0\gg 1$ and $\omega\tau_0\ll 1$, respectively. In the intermediate region a transition occurs from (5) to (1). A unified theory in which (1) and (5) appear as limiting cases has not yet been worked out.

[1] P. Parzen and L. Goldstein, Phys. Rev. **79**, 190 (1950).
[2] A. van der Ziel, J. Appl. Phys. **21**, 399 (1950).
[3] A. van der Ziel, J. Appl. Phys. **24**, 222 (1953).

Pressure-Welded p–n Junctions in Germanium

R. G. Shulman and D. M. Van Winkle
Research and Development Laboratories, Hughes Aircraft Company, Culver City, California
(Received November 24, 1952)

A RECTIFYING junction is formed in a semiconductor at a point where an abrupt change from p- to n-type material occurs.[1] One direct way to achieve such a change is to join together two crystals of different conductivity types without appreciably altering the bulk material. Because germanium expands upon solidifying, its melting temperature is lowered by the application of pressure. By localizing the maximum pressure at the surfaces to be joined, melting can be limited to the surfaces, and a mechanically strong weld will form upon solidification. A technique for making pressure-welded p–n junctions has been developed, and their electrical characteristics have been measured.

The rate of change of the melting temperature T with pressure P can be determined from the Clausius-Clapeyron equation,

$$dT/dP=T\Delta V/H. \quad (1)$$

Accepted values of the melting temperature and volume change on melting for germanium were found to be in error. They were remeasured, and a volume change[2] of 5 percent and a melting temperature of 935°C were observed. Insertion of these values, and of 110 cal/g for the heat of fusion, H, in Eq. (1) yields $dT/dP=2.4\times10^{-3}$ degree per kg/cm^2. Consideration of the practical limits of temperature control made it necessary to apply pressures of about 1000 kg/cm^2, which lower the melting point approximately two degrees. Pressures of this magnitude were achieved with moderate forces by using extremely small contact areas. The contact area was minimized by making one contact surface optically flat and the other of controlled roughness. Since only the peaks of the rough surface were in contact with the smooth surface, loads of 1 to 2 kg were sufficient for localized melting. The initial melting collapsed the peaks of the rough surface, thus bringing additional area into contact and reducing the pressure. The temperature was held constant so the melting stopped when the pressure dropped and the system stabilized itself.

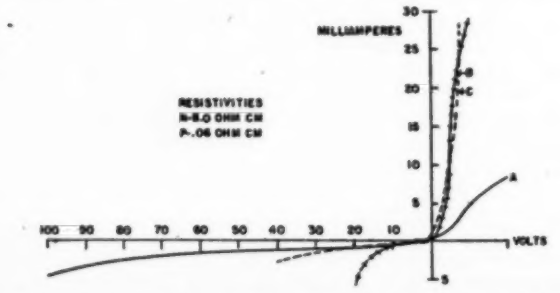


FIG. 1. Pressure-welded junction diodes direct current characteristics.

Figure 1 shows the dc characteristics of three p–n junctions formed by the pressure-weld method. All the p-type material was from the same single crystal; the n-type from another single crystal. The variations in these curves indicate that the characteristics of the junctions are largely determined by variations in the welding process. Because of the prolonged exposure to temperatures near the melting point, some solid-solid diffusion is to be expected, and the transition from p- to n-type material will not be perfectly sharp. Accordingly, the slope of the curve for logarithm of capacitance *versus* logarithm of back voltage curve should be somewhat less than $-1/2$. The observed slope was $-1/2.2$.

The lifetime of minority carriers in the imperfect welded region should be shorter than for the original undistorted crystals. One effect of the lower lifetimes is the high back saturation currents observed in these diodes. Another result is that the high frequency response of these units is better than that of rectifiers made by other methods with higher lifetimes. Figure 2 shows an oscillograph of the recovery of typical drawn, diffused and pressure-welded laboratory units. The diodes have been biased in the forward direction with a current of 8 milliamperes and a 10-μsec pulse of 5 volts applied in the backward direction.

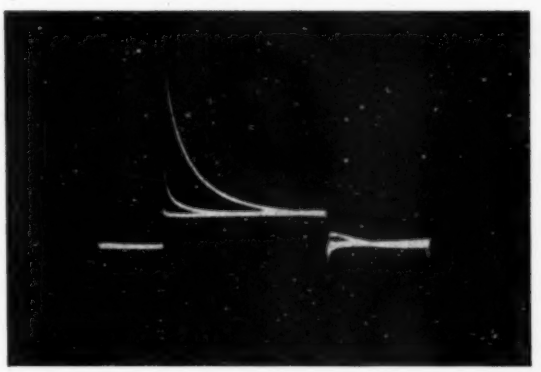


FIG. 2. Response of three diodes to a 5-volt 10-μsecond reverse pulse. Drawn diode shows longest recovery time, diffused next longest, and pressure-welded shortest.

We wish to thank H. Q. North and R. P. Johnson for many stimulating suggestions.

[1] W. Shockley, Bell System Tech. J. **28**, 435 (1949).
[2] R. C. Sangster and J. N. Carman, private communication.

A Magnetic Analog for the Interaction of Dislocation Loops

Edward W. Hart
Metallurgy Research Department, General Electric Company, Schenectady, New York
(Received October 13, 1952)

PEACH and Koehler[1] have called attention to the similarity between the force exerted on a dislocation segment by an externally applied stress field and the force exerted on a current-carrying wire by a magnetic field. The purpose of this note is to show that the magnetic analogy is even more complete for the interaction of crystallographic dislocations which are constrained to move only in the plane defined by the dislocation line and the Burger's vector when the dislocation loops and Burger's vectors

are all coplanar and all loops have the same direction of Burger's vector. This relationship would hold, for example, for the dislocation loops producing a single direction of slip in any one slip plane.

If the Burger's vector is in the x direction, and the loops lie in the xy plane, the component of stress responsible for the force is τ_{xz}.

The stress component τ_{xz}, designated hereafter simply as τ, will be computed for a dislocation loop lying in the xy plane, with Burger's vector $\mathbf{b}$ parallel to the x direction. Let $\mathbf{r}$ be the vector distance from an element $d\xi$ of the dislocation loop of the field point, let $\mathbf{u}$ be the vector displacement function, and let $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ be the usual x, y, and z unit vectors.

The displacement may be written[2,3] as

$$\mathbf{u}=(4\pi)^{-1}\{\mathbf{b}\Omega+\mathbf{b}\times\mathbf{A}+\tfrac{1}{2}(1-\nu)^{-1}b\nabla B\}. \quad (1)$$

The positive sense of the surface bounded by the loop is taken to be the direction of the z axis. With this convention, Ω is the solid angle subtended at the field point by the loop,

$$\mathbf{A}=\int r^{-1}d\xi,$$

$$B=\int(\mathbf{b}/b)\times(\mathbf{r}/r)\cdot d\xi,$$

and ν is Poisson's ratio. Now since $\mathbf{b}=b\mathbf{i}$, and since the loop lies in the xy plane, some simplifications may be made:

$$\mathbf{b}\times\mathbf{A}=b\mathbf{i}\times\mathbf{A}=b\mathbf{k}A_y,$$

$$B=\int\mathbf{i}\times\mathbf{r}\cdot r^{-1}d\xi$$

$$=-\int\mathbf{r}\cdot\mathbf{k}r^{-1}d\xi_y$$

$$=-zA_y.$$

Then

$$\mathbf{u}=(4\pi)^{-1}b\{\mathbf{i}\Omega+\mathbf{k}A_y-\tfrac{1}{2}(1-\nu)^{-1}\ zA_y\}. \quad (2)$$

Finally the stress may be written as

$$\begin{aligned}\tau&=G[(\partial u_z/\partial x)+(\partial u_x/\partial z)]\\&=(4\pi)^{-1}Gb\{(\partial\Omega/\partial z)+(\partial A_y/\partial x)-(1-\nu)^{-1}[\partial^2(zA_y)/\partial x\partial z]\}\\&=(4\pi)^{-1}Gb\{h_z-\nu(1-\nu)^{-1}(\partial A_y/\partial x)-z(1-\nu)^{-1}(\partial^2A_y/\partial x\partial z)\},\end{aligned} \quad (3)$$

where h_z is the magnitude of the z component of magnetic field which would be produced by a loop, congruent with the dislocation loop, carrying unit current in a medium of unit permeability. If the field point is in the plane of the loop ($z=0$), and if ν is zero, τ is given only in terms of h_z, and the mutual interaction of coplanar loops of the same Burger's vector direction is then seen to be given exactly in terms of an equivalent magnetic interaction. When ν is not zero, there is an additional contribution due to the second term inside the brace of Eq. (3). For applications where only the azimuthal average about the z axis is important, h_z is replaced by $[1-\frac{1}{2}\nu(1-\nu)^{-1}]h_z$ to a good approximation.

[1] M. Peach and J. S. Koehler, Phys. Rev. **80**, 436 (1950).
[2] J. M. Burgers, Proc. Roy. Acad. Sci. Amsterdam **42**, 293 (1939).
[3] F. R. N. Nabarro, Phil. Mag. **42**, 213 (1951).

Average Temperatures by Parallel-Connected Thermocouples

BERNARD E. DRIMMER*
U. S. Naval Ordnance Laboratory, White Oak, Maryland
(Received November 3, 1952)

DURING the course of an investigation at this Laboratory it became necessary to measure within one percent the mean temperature of a solid steel cylinder for about a half-hour after thermal energy had been non-uniformly applied in a short heating cycle. The use of thermocouples imbedded in the steel was indicated, but a large number of couples was necessary to attain the desired accuracy, and the problem of monitoring these couples and then reducing the data appeared formidable. The

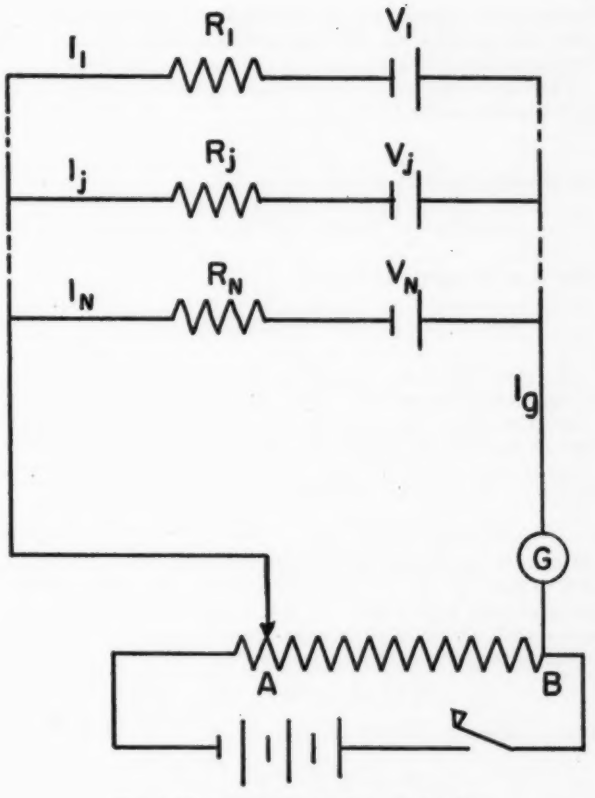


FIG. 1. Thermocouple-potentiometer circuit.

following analysis of the electric circuit resulting from the connection in parallel of all the couples shows that the observed voltage will differ from the mean voltage by less than one percent if the resistances of all legs of the circuit are matched within tolerances that are readily met. Since temperature can be considered to be linear with voltage for reasonable temperature changes, the measurement of mean temperatures to within one percent was thus achieved by the parallel connected thermocouple circuit.

If N thermocouples (voltage generators) are connected to a potentiometer as shown in Fig. 1, the current I_g, flowing through the galvanometer G, is equal to zero when the galvanometer suffers no deflection. On balance, therefore,

$$\Sigma I_j^{(1)}=0. \quad (1)$$

When the potentiometer circuit is closed, the voltage drop across the jth leg is determined by the equation $V_0-V_j=I_jR_j$, where V_0 is the observed voltage between A and B. Summing over all the legs of the couple circuit

$$V_0-\bar{V}=\frac{1}{N}\Sigma I_jR_j, \quad (2)$$

where $\bar{V}=(1/N)\ \Sigma V_j$. If all the R's were exactly equal, R_j could be factored out of the right member of Eq. (2), making that member equal to zero by Eq. (1). In this case the observed voltage is exactly equal to the mean voltage. However, resistances are never exactly matched, and it becomes necessary to determine the tolerable limits of mismatch. Let

$$\bar{R}=(1/N)\Sigma R_j;\quad \Delta R_j=R_j-\bar{R};\quad \Delta V_j=V_j-V_0.$$

The largest possible current that could flow in any leg of the couple circuit, under the given conditions, is given by the relation

$$I_{\max}=\Delta V_{\max}/R_{\min}. \quad (3)$$

(This is a conservative relation; i.e., in general the largest current

in the couple network will be smaller than I_{max} because I_{max} will flow only in that leg that simultaneously has ΔV_{max} and the minimum line resistance.)

Letting $I_1=m_1 I_{max},\cdots$, $I_j=m_j I_{max},\cdots$, at all times $|m_j|\leq 1$, and at galvanometer balance

$$\Sigma m_j=0. \tag{4}$$

Substituting these values of I_j in Eq. (2),

$$V_0-\bar{V}=\frac{I_{max}}{N}\Sigma m_j R_j=\frac{I_{max}}{N}[\bar{R}\Sigma m_j+\Sigma m_j(\Delta R_j)].$$

Hence, at galvanometer balance

$$V_0-\bar{V}=\frac{I_{max}}{N}\Sigma m_j(\Delta R_j). \tag{5}$$

If N were very large, reasonable distributions could be assigned to the values of m_j and ΔR_j so that an estimate may be made of the maximum value of the right-hand member of Eq. (5). However, N though too large to permit the monitoring of each couple individually is generally too small to permit assignment of such distributions. An absolute maximum value is therefore used:

$$|\Sigma m_j(\Delta R_j)|\leq\Sigma|m_{max}|\cdot|\Delta R_{max}|=N\cdot|\Delta R_{max}|. \tag{6}$$

(It is to be noted that $|\Sigma m_j(\Delta R_j)|$ will be equal to $N\cdot|\Delta R_{max}|$ only in the very rare case when all couples satisfy simultaneously three extreme conditions: $|m_j|=1$; $|\Delta R_j|=|\Delta R_{max}|$; the products $m_j(\Delta R_j)$ should have the same algebraic sign. A considerable safety factor still remains if one considers

$$|\Sigma m_j(\Delta R_j)|\leq(N/2)|\Delta R_{max}|).$$

Hence, from Eqs. (3), (5), and (6),

$$|V_0-\bar{V}|\leq\frac{|\Delta V_{max}|\cdot|\Delta R_{max}|}{R_{min}}. \tag{7}$$

Let $|P|\cdot\bar{V}$ be the maximum permissible error introduced by the mismatch of resistances, i.e.,

$$|V_0-\bar{V}|\leq|P|\cdot\bar{V}. \tag{8}$$

Equation (7) shows that Eq. (8) is satisfied when

$$|\Delta V_{max}|\cdot\frac{|\Delta R_{max}|}{R_{min}}\leq|P|\bar{V},$$

or

$$\frac{|\Delta V_{max}|}{\bar{V}}\cdot\frac{|\Delta R_{max}|}{R_{min}}\leq|P|. \tag{9}$$

For temperature changes of the order of 50°C or less, negligible errors are introduced by assuming that $V\propto T$, where T is the difference in temperatures between the thermocouple junctions. Thus,

$$\frac{|\Delta T_{max}|}{\bar{T}}\cdot\frac{|\Delta R_{max}|}{R_{min}}\leq|P|. \tag{10}$$

Equation (10) shows that $|P|$ decreases with increasing $\bar{T}$ and R_{min}. The former is assigned by experimental conditions while the minimum line resistance can be manipulated within wide limits. However, the limiting value of R_{min} is determined by the fact that the sensitivity of the potentiometer decreases with increasing resistance of the circuit whose voltage is under measurement.[2] Therefore, R_{min} should be kept as low as is consistent with both Eq. (10) and the sensitivity of the galvanometer. For any given apparatus, where the ratio $|\Delta R_{max}|/R_{min}$ is constant, $|P|$ approaches zero as the material being measured becomes thermally homogeneous, i.e., $|\Delta T_{max}|$ approaches zero. Thus, the observed temperature differs from the mean temperature by smaller and smaller amounts as the material tends to thermal homogeneity, which is as it should be.

To set up an apparatus where $|P|$ is to be held to some maximum value, it is necessary to estimate the ratio $|\Delta T_{max}|/\bar{T}$. [It is necessary, in placing the couples within the material under measurement, that they be so positioned that each has "cognizance" over an equal volume (actually, heat capacity) of the material under test.] In the survey tests for our experiments individual thermocouples (as many as twenty in the steel bar) were monitored. It was found that beyond two minutes after the energy was fed to the steel bar $|\Delta V_{max}|/\bar{V}\leq 0.3$. Using a factor of two to be on the safe side, it was considered that beyond that moment the ratio was less than 0.6. (Equilibrium temperatures at zero time were then obtained by extrapolation.) Thus, if $|P|$ were 0.01, $|\Delta R_{max}|/R_{min}\leq 0.017$. Since R_{min} was approximately three ohms, a $|\Delta R_{max}|$ of 0.051 ohm gave assurance that the observed temperature differed from the mean temperature by less than one percent.

A detailed analysis of the components of R_j showed that the hot junctions of the couples may be soldered to the steel bar without forcing $|P|$ to a higher value than 0.01. For a steel bar 13 cm long and three-cm diameter the maximum uncertainty in measuring R_j was 0.004 ohm, allowing 0.047 ohm for mismatch and for drift in R_j with time over several months.

In the derivation of Eq. (10) two highly conservative relations were used—Eq. (3) and (6). If one uses the largest current actually flowing in the network instead of I_{max} and also the actual value of $|\Sigma m_j(\Delta R_j)|$ instead of its maximum possible value of $N\cdot|\Delta R_{max}|$, it can be shown that the ratio $|\Delta R_{max}|/R_{min}$ can be larger than $|P|\cdot\bar{V}/|\Delta V_{max}|$ by a factor of at least two except under exceedingly rare conditions. However, the requirement of making $|\Delta R_{max}|\leq 0.051$ ohm in the steel bar experiment was readily met, so that the data obtained were considered better than the one percent that was claimed.

* Present address: U. S. Atomic Energy Commission, 19 Street and Constitution Avenue, N. W., Washington, D. C.

[1] In the following discussion $\Sigma=\sum_{j=1}^{N}$.

[2] See, for example, G. P. Harnwell, *Principles of Electricity and Electromagnetism* (McGraw-Hill Book Company, Inc., New York, 1938), p. 328.

The Hall Effect in Titanium

GEORGIANA W. SCOVIL
Bryn Mawr College, Bryn Mawr, Pennsylvania
(Received November 7, 1952)

PRELIMINARY measurements of the Hall coefficient for titanium give a value of $+(2.8\pm0.9)\times10^{-11}$ VM³/AWb. A dc method was employed for these measurements, which were made at a temperature of about 100°C and in a magnetic field of 0.40 Wb/M². The Hall voltage was detected by a galvanometer having a sensitivity of 5×10^{-8} V/mm. Measurements made on a copper sample agreed with the established value[1] of -5.5×10^{-11} VM³/AWb, and were therefore a check on the calibration of the apparatus.

The anomalous Hall effect indicated by the positive coefficient may be most simply explained by assuming that holes as well as electrons contribute to conduction and, therefore, that there is an overlapping of the energy bands. If a free electron model is assumed, the order of magnitude of the densities of the holes and electrons can be obtained from the following expressions[2] for the conductivity σ, and the Hall coefficient R:

$$\sigma=e(\mu_n n+\mu_p p), \tag{1}$$

$$R=(-nb^2+p)/(nb+p)^2 e, \tag{2}$$

where n and p are the densities of the electrons and holes, respectively, μ_n and μ_p the corresponding mobilities, b the ratio of μ_n/μ_p, and e is the electronic charge. The value of the conductivity[3] used is $1.2\times10^6/\Omega$M. Since the holes are presumably at the top of a d band and the electrons at the bottom of an s band, it has been assumed that the ratio of the effective mass of the holes to that of the electrons is 16 and that the mobility of the electrons is not far different from the mobility of electrons in copper, which is 3.5×10^{-3} M²/VS. The latter is the product of the Hall coefficient and the conductivity. The densities calculated are 0.3×10^{27} electrons/M³ and 7.5×10^{27} holes/M³.

From this density of electrons in the 4s band, and estimation can be made of ϵ', the depth of the bottom of this band below the Fermi level, for

$$n=[4\pi(2m)^{\frac{3}{2}}/h^3]\int_0^{\epsilon'}\sqrt{\epsilon}\,d\epsilon, \tag{3}$$

where m is the mass of the electron, h is Planck's constant, and ϵ is the energy of a level in the band measured from the bottom of the band. The solution of Eq. (3) gives ϵ' as 0.12 electron volt.

This work is preliminary to a study of the variation of the Hall coefficient with temperature which is planned in connection with an investigation of the transition point of titanium at 882°C. It is part of a study of the physical properties of titanium being carried on in this laboratory.[3]

[1] *International Critical Tables* (McGraw-Hill Book Company, Inc., New York, 1933), Vol. 6, p. 416.
[2] William Shockley, *Electrons and Holes in Semiconductors* (D. Van Nostrand Company, Inc., New York, 1950), pp. 216–217.
[3] Walter C. Michels and Sally Wilford, J. Appl. Phys. **20**, 1223 (1949).

Space Charge Limited Currents between Inclined Plane Electrodes. Approximate Solutions

HENRY F. IVEY
Skiatron Electronics and Television Corporation, New York, New York
(Received November 12, 1952)

THE problem of space charge limited current flow between inclined plane electrodes has recently been discussed in detail by the writer[1] using a method suggested by Walker.[2] It was shown there that the cathode current density corresponding to complete space charge limitation for electrons emitted with no initial velocity from a point at a distance r_0 from the apex of the electrode system (see Fig. 1) is given by

$$J_0=2.330V_a^{\frac{3}{2}}r_0^{-2}F(\theta)\mu a/cm^2, \tag{1}$$

where θ is the angle between the electrodes and $F(\theta)$, the "perveance function," is expressed as a series expansion in terms of θ.

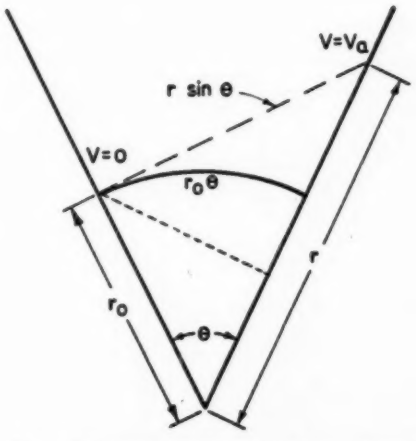


FIG. 1. Electrode arrangement and nomenclature for inclined plane electrodes.

The current density collected by the anode at a point at a distance r from the apex of the electrodes for complete space charge limitation is given by

$$J_a=2.330V_a^{\frac{3}{2}}r^{-2}h(\theta)\cdot F(\theta)\mu a/cm^2, \tag{2}$$

where $h(\theta)$, the "trajectory magnification factor," is also expressed as an infinite series in terms of θ.

The series expressions for $h(\theta)$ and $F(\theta)$ are quite awkward for computation, and it would be very desirable if simple approximate expressions could be derived, particularly for small values of θ. It has been found possible to obtain such expressions from elementary considerations based on the well-known result for parallel plane electrodes,

$$J=2.330V_a^{\frac{3}{2}}d^{-2}\mu a/cm^2. \tag{3}$$

Here d is the distance between the electrodes.

For purposes of approximation in the case of inclined electrodes, for the point on the cathode which is at a distance r_0 from the apex, one could take for the equivalent interelectrode spacing for parallel plane electrodes either the length of the perpendicular to the cathode at that point (the dashed line of the figure) or the length of the perpendicular to the anode which passes through the point considered (the dotted line of Fig. 1). Since it is known that the magnification factor is always greater than unity ($r>r_0$) the first is the better choice. The equivalent interelectrode spacing d is then given by $r_0\tan\theta$ or by $r\sin\theta$.

If the first of these expressions is substituted in Eq. (3), and the result is compared with Eq. (1), one obtains

$$F(\theta)\approx\cot^2\theta, \tag{4}$$

while if the second expression is substituted in Eq. (3) and the result is compared with Eq. (2), one obtains

$$h(\theta)\cdot F(\theta)\approx\csc^2\theta. \tag{5}$$

Numerical comparison with the values calculated earlier from the series expressions shows that of these two results, Eq. (5) is much the better approximation. The accuracy of this approximation is shown in the third column of the accompanying table; the error is about 1 percent at 20° and 5 percent at 45°.

TABLE I. Accuracy of approximations for space charge limited current flow between inclined plane electrodes.

θ deg.	θ rad.	$\frac{\csc^2\theta}{h(\theta)F(\theta)}$	$\frac{\theta^{-2}}{F(\theta)}$	$\frac{(\theta/\sin\theta)^2}{h(\theta)}$
0	0	1.0000	1.0000	1.0000
5.73	0.1	1.0007	1.0004	1.0003
11.46	0.2	1.0030	1.0016	1.0014
17.19	0.3	1.0066	1.0036	1.0030
22.92	0.4	1.0119	1.0065	1.0053
28.65	0.5	1.0185	1.0102	1.0082
34.38	0.6	1.0267	1.0149	1.0116
40.11	0.7	1.0368	1.0206	1.0159
45.84	0.8	1.0485	1.0273	1.0206
51.57	0.9	1.0619	1.0351	1.0259
57.30	1.0	1.0772	1.0442	1.0316
63.03	1.1	1.0953	1.0547	1.0385
68.76	1.2	1.1150	1.0673	1.0449
74.49	1.3	1.136	1.0805	1.0518
80.22	1.4	1.163	1.0963	1.0612
85.95	1.5	1.193	1.114	1.0703
90.00	$\pi/2$	1.217	1.129	1.0779

If only the first two terms in the series expression for $F(\theta)$ are retained,[1] the result is

$$F(\theta)=\theta^{-2}(1+0.02667\theta^2)^{-1}. \tag{6}$$

This expression is itself not very convenient for calculation. It is found, however, that the simple approximation $F(\theta)\approx\theta^{-2}$ is surprisingly good, as shown by the fourth column of the Table I; the error is about 1 percent at 28° and 5 percent at 60°. It may be noted that this approximation corresponds to taking the equivalent parallel electrode spacing, d in Eq. (3), as $r_0\theta$, that is, the distance between the electrodes as measured along the circular arc passing through the point at a distance r_0 from the apex.

Since we have now established the approximations $F(\theta)\approx\theta^{-2}$ and $h(\theta)\cdot F(\theta)\approx\csc^2\theta$, it follows that $h(\theta)\approx(\theta/\sin\theta)^2$. The accuracy of this expression is shown in the last column of Table I; the error is about 1 percent at 30° and 5 percent at 75°.

[1] H. F. Ivey, J. Appl. Phys. **23**, 240 (1952).
[2] G. B. Walker, Proc. Phys. Soc. (London) **63(B)**, 1017 (1950).

Evaporated Point Contact Rectifiers

E. G. Roka, C. H. Jackson, and R. P. Ulrich
General Research Laboratory, Minneapolis-Honeywell Regulator Company, Minneapolis, Minnesota
(Received September 8, 1952)

IN the ordinary germanium point contact rectifier we have replaced the "cat whisker" by a very small point contact of evaporated silver. The influence of this alteration on the characteristics has been studied. Important differences have been found between characteristics with tungsten cat whiskers and evaporated silver contacts, including significantly better rectifying properties. On the other hand, when a power treatment is given using a cat whisker and the cat whisker then replaced with evaporated silver, the characteristics in the back direction are practically the same up to 30 volts.

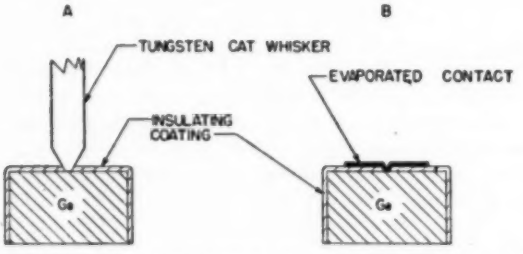


FIG. 1. Preparation of evaporated point contact germanium rectifiers. (A) The germanium surface is covered with an insulating layer in which a hole of about one mil diameter is punched with a tungsten "cat whisker." (B) After removal of the "cat whisker" a silver contact is evaporated in vacuum. (The figure is not drawn to scale.)

The structure of the evaporated point contact rectifiers is shown in Fig. 1. The germanium surface was covered with a thin insulating layer. A hole of about one mil diameter was punched in the insulating layer with a tungsten cat whisker. After removal of the cat whisker, the surface was covered with silver by evaporation in vacuum. The dc characteristics were measured (a) before coating, (b) with cat whisker in place after coating and punching, and (c) with evaporated silver. In cases (b) and (c) the same point on the germanium surface was used. These measurements were made without any power treatment.

Figure 2A shows average characteristics of 12 units using a cat whisker before coating. Figure 2B shows the average characteristic when the cat whisker was placed on the germanium through a hole in the insulating film. A force of about 10 grams on the cat whisker was sufficient in both cases. Characteristics in the two cases show no significant difference.

The average characteristic for evaporated silver electrodes is shown in Fig. 2C. The back resistance is considerably higher than for a cat whisker. The forward current at one volt was found to

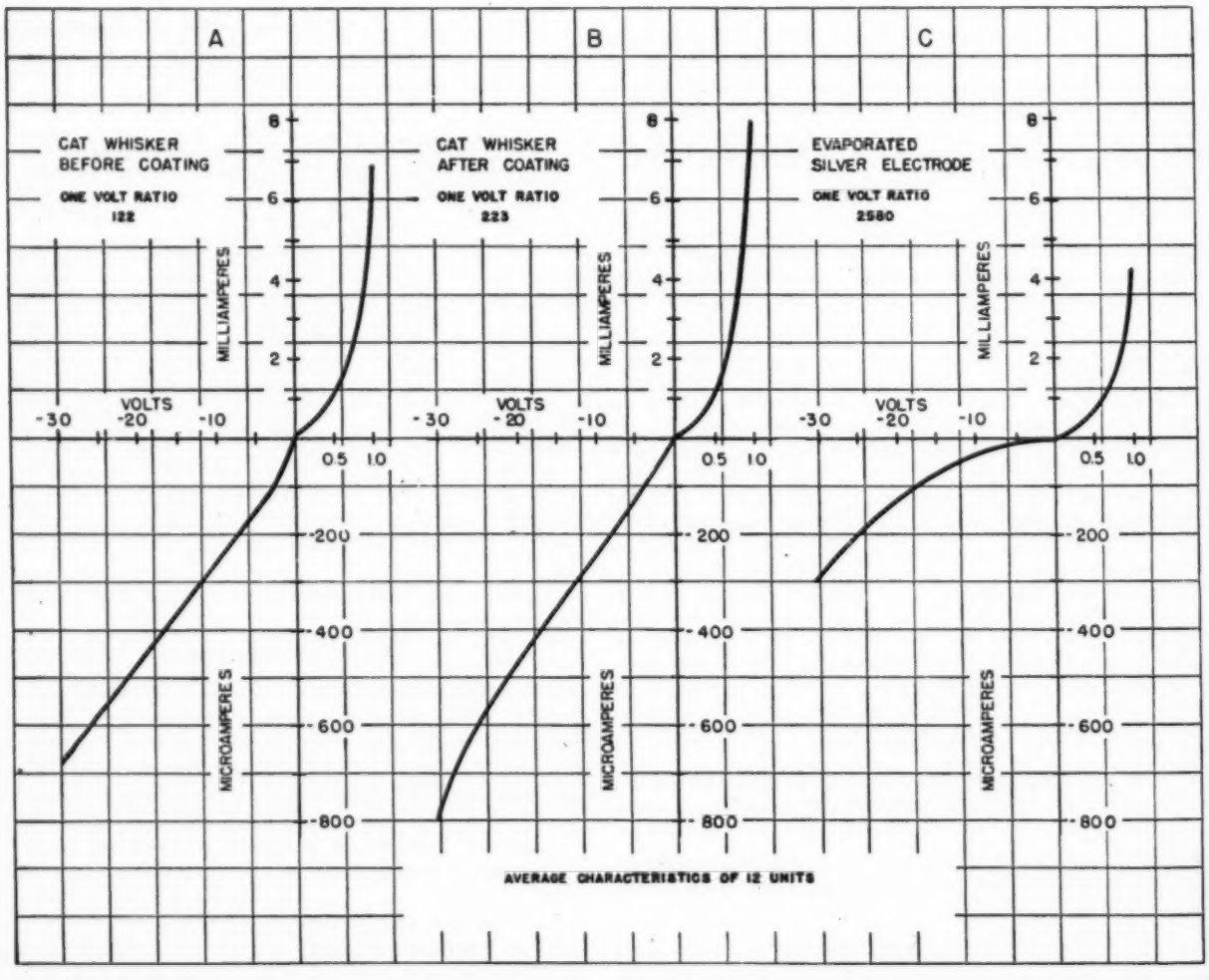


FIG. 2. Characteristics without power treatment. (A) Tungsten "cat whisker" on free germanium surface. (B) Cat whisker on germanium when the germanium surface is covered with a thin insulating layer. (C) The cat whisker was replaced by a silver deposit evaporated in vacuum. The curves for the cat whisker with and without insulating layer are practically the same. The evaporated silver electrode shows considerably better rectifying properties than the cat whisker.

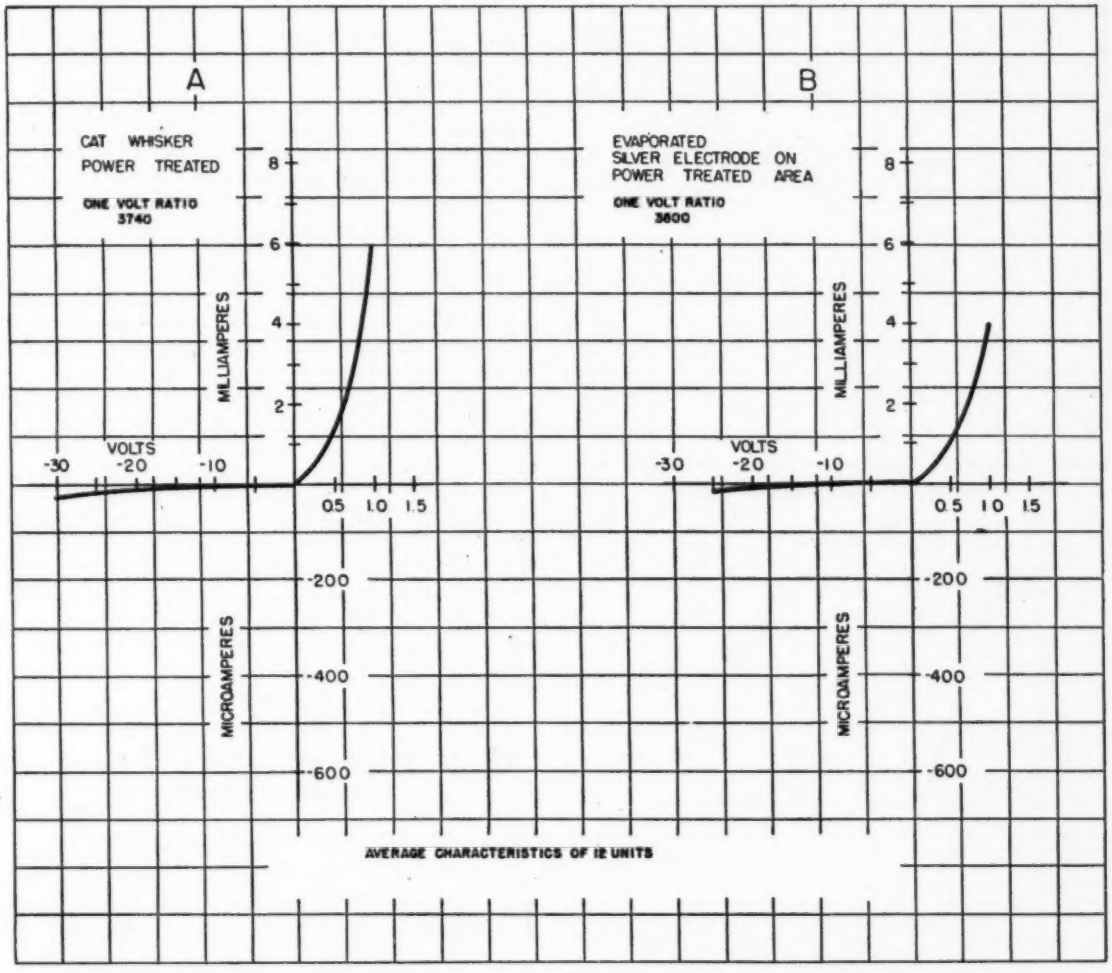


FIG. 3. Evaporated point contact on power treated area. (A) Average characteristic of "cat whisker" units after routine power treatment. (The characteristic before power treatment is shown in Fig. 2B.) (B) Average characteristics for evaporated silver electrode substituted for power treated cat whisher. The surface of the germanium under the cat whisker is permanently changed by the power treatment, and the substitution of evaporated silver contact for the cat whisker produces practically no change in back characteristics up to 30 volts in contrast to the case when no power treatment was applied.

decrease by a factor of about two; however, the rectification ratio at one volt for silver was better by about one order than for the cat whisker. The rectification ratio averaged 2600. The slope of the characteristic curve at low voltage was smaller than the theoretical value of $\alpha=39$-1 volts for both the silver electrode and the cat whisker, but slightly better for silver than for the cat whisker.

In a second group of measurements the effect of power treatment was studied. The power treatment was applied with a tungsten cat whisker before making the evaporated silver contact. Figure 3A shows the average characteristic of 12 units which received a routine power treatment with the cat whisker through the insulating coating. The power treatment consisted of short ac pulses at about 30 volts, lasting from 4/10 to 6/10 seconds. When the cat whisker was replaced by an evaporated silver electrode after power treatment, the characteristics in Fig. 3B were obtained.

In contrast to the previous measurements without power treatment, the back resistance remained practically unchanged when silver was substituted for the cat whisker. The power treatment seemed to produce a permanent effect on the germanium. The effect of the power treatment persists in the sense that the first metal contact can be replaced with another without producing a significant change in the back characteristic up to 30 volts. For the case of power treatment with a cat whisker, the slope α in the low voltage range was slightly over 30, while the for silver electrode on the power treated area, the value of α was slightly lower.

We wish to express our gratitude to Dr. W. H. Kliever for his assistance in making this publication possible. We are also indebted for assistance in various aspects of this work to the late R. T. Squier, to Dr. V. W. Bearinger, and F. M. Exner.

Radiation Damage of Beta-Brass*

R. R. EGGLESTON AND F. E. BOWMAN
North American Aviation, Inc., Atomic Energy Research Department, Downey, California
(Received November 7, 1952)

IN the field of superlattice alloys there is a group in which the disordered state cannot be obtained at room temperature by quenching from above the critical temperature. The phenomena is apparently associated with a low activation energy for the ordering process and the resulting high relaxation rates at temperatures well below the critical. One of the most typical alloys of this type is beta-brass. Evidence that the relaxation rate remains high even at room temperature is provided by the fact that disordering cannot be observed following cold working.

As a part of an investigation of the effects of cyclotron irradiation on the order-disorder transformation, it was decided to include beta-brass because of its rather unique characteristics.

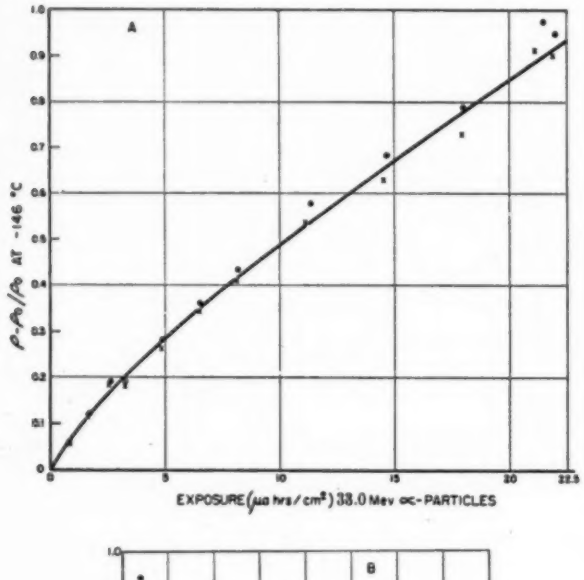


FIG. 1. A. Change in resistivity of beta-brass on exposure to 33.0-Mev alpha-particle bombardment at temperatures below −100°C. B. Change in resistivity on pulse annealing after radiation damage.

It has been demonstrated that high energy charged particles are capable of disordering the ordered copper-gold alloys if they are bombarded below their critical temperature.[1] If beta-brass was to be similarly bombarded at a sufficiently low temperature and investigated at this temperature evidence of disordering should become apparent. Once disordered it was hoped that further information relative to the kinetics of ordering could be made available. We wish to present here the results of the irradiation studies.

Two beta-brass foil specimens, 5 mil thick, were irradiated with 33.0-Mev alpha-particles on the 60-inch cyclotron of the University of California at Berkeley. The specimens were maintained at low temperatures during the bombardment by blowing helium cooled to liquid nitrogen temperatures past them; the temperature was monitored during the irradiation and was never allowed to exceed −100°C. At convenient intervals during the irradiation the bombardment was stopped and the specimen resistance was measured at the base temperature of −146°C. The results of these measurements are given in Fig. 1A as a function of alpha-particle exposure.

After an irradiation of 23 μahr/cm², the specimens were annealed by "pulsing" them for five minutes at successively higher temperatures. The specimens, however, were returned to the base temperature for a resistance measurement between each five-minute pulse before being subjected to the next higher temperature pulse. The results of this annealing schedule or the specimens are shown in Fig. 1B.

The interesting features of the experimental curves are the following: (1) The large percentage increase in resistivity that was induced in the specimen. Under similar conditions a pure metal, e.g., copper, the change would have been between a 20 percent and 25 percent increase, so it seems unlikely that the mechanism causing the large observed increase in the beta-brass would be similar to that causing the increase in copper. (2) The annealing curves after irradiation show that the induced damage anneals out very rapidly at relatively low temperatures. After a pulse anneal at room temperature the damage has largely annealed, demonstrating the importance of maintaining a low temperature during the experiment.

In view of the above discussion it seems not at all unlikely that the observed changes noted in the experiment were due to the alloy being transformed to the disordered state by the alpha-particle bombardment.

The authors wish to take this opportunity to thank F. E. Fillmore and the Berkeley group of North American Aviation, Inc., for their valuable help in performing the bombardment and subsequent annealing.

* This report is based on studies conducted for the U. S. Atomic Energy Commission under contract AT-11-1-GEN-8.

[1] S. Siegel, Phys. Rev. **75**, 1823 (1949); see also Martin, Austerman, Eggleston, McGee, and Tarpinian, Phys. Rev. **81**, 664 (1951).

On the Theory of Prediction of Nonstationary Stochastic Processes

GLENN W. PRESTON
Philco Corporation, Research Division, Philadelphia, Pennsylvania
(Received October 6, 1952)

IN a recent issue of the *Journal of Applied Physics*, R. C. Davis[1] derives the optimum linear predictor of the form

$$S^*(T+\Delta)=\int_0^T K(t)[S(t)+N(t)]dt,$$

in which $S(t)$ signifies signal and $N(t)$ noise, both of which may be nonstationary.

His result for the case $ES(t)\equiv 0$ is

$$K(t)=\Sigma_i\lambda_i\phi_i(t)\int_0^T du r_{SN}(T+\Delta, u)\phi_i(u), \qquad (1)$$

in which ϕ_i are the eigenfunctions and λ_i the corresponding eigenvalues of the integral equation

$$\phi(s)=\lambda\int_0^T R_{SN}(s, t)\phi(t); \qquad (2)$$

also

$$r_{SN}(s, t)=E\{[S(s)-m(s)][S(t)-m(t)+N(t)]\},$$

and $R_{SN}(s, t)=E\{[S(s)-m(s)+N(s)][S(t)-m(t)+N(t)]\}$ by definition.

For other than Markoff processes, the author was unable to make contact with Wiener's theory of prediction, owing to the fact that the theory of integral equations which he employs does not apply when the range of integration is infinite. The following development appears to show that this fact presents no difficulty.

We replace $K(t)$ by $q(T-t)$ which can then be identified with the memory function of the predicator filter. Then,

$$\begin{aligned}\int_0^T dt q(T-t)R_{SN}(u, t)&=\Sigma\lambda_i\int_0^T dt R_{SN}(u, t)\phi_i(t)\\ &\quad\cdot\int_0^T ds r_{SN}(T+\Delta, s)\phi_i(s)\\ &=\Sigma\phi_i(u)\int_0^T ds r_{SN}(T+\Delta, s)\phi_i(s)\\ &=r_{SN}(T+\Delta, u),\end{aligned}$$

(recalling that the ϕ_i are an orthonormal complete set in the interval 0 to T).

Making the change of variables $x=T-t$, $y=T-u$, this becomes

$$r_{SN}(T+\Delta, T-y)=\int_0^T dx q(x) R_{SN}(T-y, T-x). \quad (3)$$

The relationship between Davis' result, Eq. (1), and Wiener's theory[2] as extended by Zadeh and Ragazzini[3] is perhaps clearer when put into its equivalent form Eq. (3).

If the processes are stationary, $r_{SN}(u, v)=r_{SN}(u-v)$ and $R_{SN}(u, v)=R_{SN}(u-v)$, and if T becomes infinite,

$$r_{SN}(\Delta+y)=\int_0^\infty dx q(x) R_{SN}(x-y).$$

Except for a change of notation, this is Wiener's integral equation for the optimum weight function $q(t)$ giving a prediction time Δ.

[1] R. C. Davis, J. Appl. Phys. **23**, 1047 (1952).
[2] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series* (John Wiley and Sons, Inc., New York, 1949).
[3] Lofti A. Zadeh and John R. Ragazzini, J. Appl. Phys. **21**, 645 (1950).

Evidence for Collapse of Lattice Vacancy Aggregates to Form Dislocation Rings*

R. W. Hoffman, F. J. Anders, and E. C. Crittenden, Jr.
Case Institute of Technology, Cleveland, Ohio
(Received November 7, 1952)

Seitz[1] has proposed that, when the density of lattice vacancies exceeds the equilibrium value, the vacancies will coalesce into progressivley larger groups until they finally organize into sheets and collapse to form dislocation rings. Preliminary evidence, based on indirect measurement of the volume change occurring in evaporated metal films during annealing at progressively higher temperatures, indicates the existence of the above mechanism.

The volume change is deduced from observation of the tensile stress in evaporated Ni films. The metal is evaporated onto 0.002 $\times\frac{1}{4}\times$3-in. strips of mica as a substrate. These are solidly clamped against a Pb backing to ensure thermal contact. Four successive specimens of films were formed at 75°C, at a constant rate of 60 angstroms/sec, to thicknesses of about 2000 angstroms in a residual pressure of 4×10^{-5} mm of Hg. The curvature of the substrate strip after release from the clamps gives a quantitative measure of the stress in the metal film.[2] The curvature was measured at 25°C. Figure 1 shows the stresses observed in the Ni films after annealing for 30 minute intervals. The films are found to be in a high initial tension. Simple differential expansion of the Ni and mica contributes a stress of only 0.5×10^9 dynes/cm^2 in changing the temperature from 75°C to 25°C. Figure 1 shows

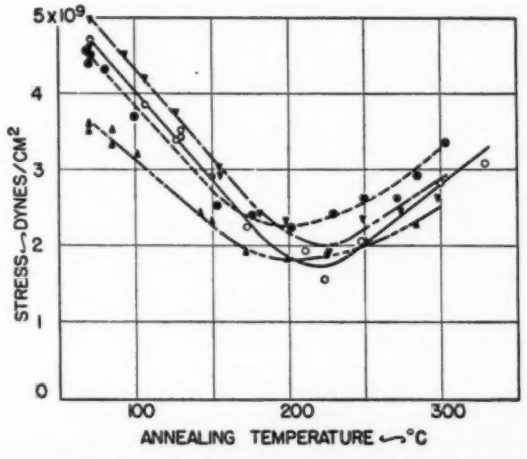


FIG. 1. Stress in Ni films *vs* annealing temperature. The stress was measured at 25°C, after annealing for 30 min at the temperatures indicated.

that on annealing the stress first falls and then later rises. The electrical resistivity continuously falls with annealing through the same temperature range, as shown in Fig. 2. If one assumes

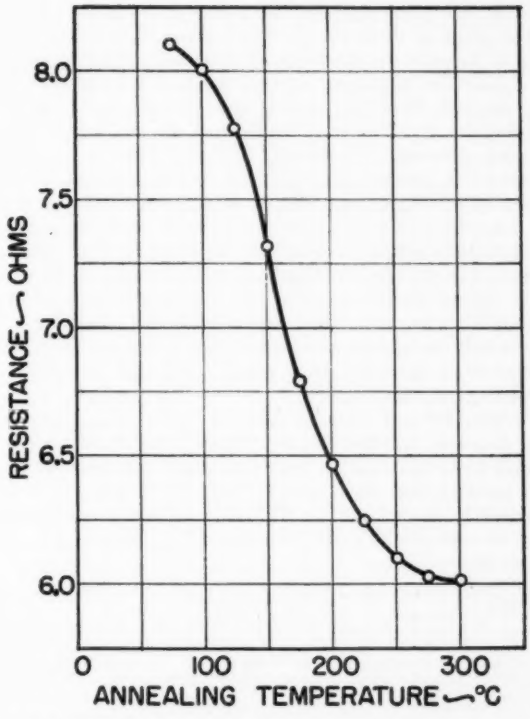


FIG. 2. Electrical resistance of Ni film *vs* annealing temperature. The resistance was measured at 50°C after annealing for 30 min at each successive temperature.

that the deposition process has caused crystals of the polycrystalline film to contain lattice vacancies, loss of these vacancies would cause the crystal to expand, i.e., if a vacancy is filled, the neighboring atoms relax outward. Coalescence of single vacancies into pairs or larger groups will also give expansion, since the number of neighboring atoms per vacancy is reduced by this coalescence. If Seitz's[1] model is followed, and the vacancies organize to sheets which then collapse to form dislocation rings, the expansion should be followed by a large shrinkage of the crystals.

Measurements of hardness, x-ray diffraction line breadth, and electrical resistivity on cold-worked Ni reported by Wilson and Thomassen[3] show that these three properties drop rapidly for 1 hour annealing in the vicinity of 600°C. This process presumably corresponds to the loss of dislocations. A large drop in resistivity is also reported by Wilson and Thomassen in a region centering at about 150°C, whereas the hardness is essentially unaffected. This indicates that the resistance change in this lower temperature range is due to imperfections other than dislocations. The drop in resistivity observed in evaporated Ni films, as shown in Fig. 2 for 30 minutes annealing, has its center at 160°C and is thus presumably the same phenomenon as reported by Wilson and Thomassen. The difference in time of annealing would account for the slight temperature difference.

Referring to Fig. 1, the decrease in tensile stress on annealing represents an expansion such as would be provided by coalescence of vacancies. The reversal at higher temperature to give an increase in tension represents a contraction such as would be provided by collapse of sheets of vacancies to form dislocation rings. Thus, we believe that the imperfections responsible are lattice vacancies and that some process very like that suggested by Seitz must be occurring.

The reversal of stress change shown in Fig. 1 would seem to suggest that the mica substrate may be changing size irreversibly.

This has been checked and there seems to be no effect. With regard to reproducibility the mica strips have all been cut in the same crystallographic direction from the mica sheets.

That evaporated films have a bad reputation as regards possessing the true properties of bulk metal is recognized. However, careful study of these films indicates them to be normal metallic Ni. The vacuum and condensation rate employed allow a possible gas impurity of 10 percent, if every incident residual gas molecule were trapped. However, the effects observed are not sensitive to gas pressure in this range. The Ni used was electroplated onto a tungsten wire source from salts of 99.8 percent purity. The remaining 0.2 percent is principally cobalt. Surface scattering effects in the evaporated films at thicknesses small compared to the conduction electron mean free path have been carefully studied. These effects are negligible at the large thicknesses used here so that the electrical and magnetic properties are normal. From measurements on resistivity and temperature coefficient of resistance for very thin specimens (5–100 angstroms), it is known that the specimens are slabs, flat within less than about 5 angstroms[4] except for possible steps in the mica backing, which would have no effect.

Electron diffraction studies have been made by Halteman[5] on films prepared in the same equipment. He has found the line breadth to be entirely that resulting from microstress rather than from particle size. This indicates that, although the films are finely divided polycrystalline Ni, the principal effects observed are those occurring within the crystals rather than at the crystal boundaries.

* Supported by the Office of Naval Research.
[1] F. Seitz, Advances in Physics 1, 43 (1952).
[2] E. C. Crittenden, Jr., and R. W. Hoffman, Phys. Rev. 78, 349 (1950).
[3] J. E. Wilson and L. Thomassen, Trans. Am. Soc. Metals 22, 769 (1934).
[4] E. C. Crittenden, Jr., and R. W. Hoffman, Office of Naval Research, Washington Conference on Magnetism, September 2–6, 1952. To appear in Revs. Modern Phys.
[5] E. K. Halteman, J. Appl. Phys. 23, 150 (1952).

Theory of Wave-Guide-Fed Slots Radiating into Parallel Plate Regions

L. Lewin
Standard Telecommunication Laboratories Limited, Enfield, Middlesex, England
(Received October 27, 1952)

In a paper of the above title H. Gruenberg[1] introduces the following integrals:

$$Q(0)=\int_0^{\frac{1}{2}\pi}\frac{\cos^2(\frac{1}{2}\pi\sin\theta)}{\cos^2\theta}\,d\theta,$$

$$Q(\xi)=\int_0^{\frac{1}{2}\pi}\frac{\cos^2(\frac{1}{2}\pi[1-\xi^2]^{\frac{1}{2}}\sin\theta)}{1-(1-\xi^2)\sin^2\theta}\,d\theta.$$

$Q(0)$ is evaluated by numerical integration.

It is the purpose of this note to give an alternative simple expression for $Q(\xi)$ and to complete the integration of $Q(0)$ in terms of tabulated functions.

The expression for $Q(\xi)$ can be put in the form

$$\begin{aligned}Q(\xi)&=\frac{1}{4}\int_0^{\frac{1}{2}\pi}\left\{\frac{1+\cos(\pi[1-\xi^2]^{\frac{1}{2}}\sin\theta)}{1+(1-\xi^2)^{\frac{1}{2}}\sin\theta}+\frac{1+\cos(\pi[1-\xi^2]^{\frac{1}{2}}\sin\theta)}{1-(1-\xi^2)^{\frac{1}{2}}\sin\theta}\right\}d\theta\\&=-\frac{1}{4}\int_0^{\frac{1}{2}\pi}\left\{\frac{\cos[x(1+(1-\xi^2)^{\frac{1}{2}}\sin\theta)]}{1+(1-\xi^2)^{\frac{1}{2}}\sin\theta}+\frac{\cos[x(1-(1-\xi^2)^{\frac{1}{2}}\sin\theta)]}{1-(1-\xi^2)^{\frac{1}{2}}\sin\theta}\right\}\Bigg|_{x=0}^{x=\pi}d\theta\\&=\frac{1}{2}\int_0^{\pi}\int_0^{\frac{1}{2}\pi}\sin x\cos[x(1-\xi^2)^{\frac{1}{2}}\sin\theta]\,d\theta\,dx\\&=\frac{\pi}{4}\int_0^{\pi}\sin x J_0[x(1-\xi^2)^{\frac{1}{2}}]\,dx.\end{aligned}\tag{1}$$

This is the expression to which we referred. To evaluate $Q(0)$, integrate twice by parts

$$\int\sin xJ_0(ax)\,dx=x[\sin xJ_0(ax)-a\cos xJ_1(ax)]-(1-a^2)\int x\cos xJ_0(ax)\,dx.$$

Taking $a=(1-\xi^2)^{\frac{1}{2}}$ and the limits of x from 0 to π, we find for $\xi=0$, $a=1$, the value

$$Q(0)=\frac{\pi^3}{4}J_1(\pi)=0.7022.$$

Using the result[2]

$$J_0[x(1-\xi^2)^{\frac{1}{2}}]=\sum_0^{\infty}\frac{Jm(x)}{\lfloor m}(\tfrac{1}{2}x\xi^2)^m$$

and

$$\int x^mJ_m(x)\sin x\,dx=\frac{x^{m+1}}{2_{m+1}}[\sin xJ_m(x)-\cos xJ_{m+1}(x)],$$

the expression for $Q(\xi)$ can be expanded into

$$Q(\xi)=\frac{\pi^3}{4}\sum_0^{\infty}\frac{(\frac{1}{2}\pi\xi^2)^m}{\lfloor m}\frac{J_{m+1}(\pi)}{2_{m+1}}$$

which is suitable for numerical computation for small values of ξ. The value of $\int_0^1Q(\xi)d\xi$ can be obtained readily from (1) by putting $\xi=\cos\theta$ and using the integral[3]

$$\int_0^{\frac{1}{2}\pi}J_0(x\sin\theta)\sin\theta d\theta=(\tfrac{1}{2}\pi)^{\frac{1}{2}}J_{\frac{1}{2}}(x)x^{-\frac{1}{2}}=\sin x/x.$$

Hence

$$\int_0^1Q(\xi)d\xi=\frac{\pi}{4}\int_0^{\pi}\frac{\sin^2x}{x}\,dx,$$

which is equivalent to Gruenberg's equation (27).

[1] H. Gruenberg, J. Appl. Phys. 23, 733 (1952).
[2] Jahnke and Ende, *Table of Functions*, p. 145.
[3] Watson, *Theory of Bessel Functions*, p. 373.

On the Zr-Hf System*

R. B. Russell
Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received October 27, 1952)

Recently Fast,[1] in reviewing an earlier investigation by Duwez[2] on the Zr-Hf system, estimated that Duwez's hafnium was actually only about 81 atomic percent pure. This appraisal was based on Fast's determination of the transformation temperature of a 94 atomic percent Hf alloy as 1950±100°C instead of 1310±10°C as found by Duwez for his own metal; moreover, Fast's doubts were further based on Duwez's somewhat larger hafnium lattice parameters.

In August, 1949, Dr. A. R. Kaufmann of this laboratory received from the Philips Company in Eindhoven, Holland, three iodide hafnium-zirconium alloys with a Philips analysis of 62, 97, and 100 atomic percent hafnium.[3]

Lattice parameters of these and of those reported in the literature[1,2,4,5] are given in Table I. It is seen that if certain assumptions are made about the kind of units used in reporting lattice parameters, many discrepancies effectively disappear.

In 1947 the Bragg Committee[6] recommended that when lattice parameters are reported, the units used should be uniquely specified. It is suggested that some of the disagreement about the Zr-Hf system can be resolved if the Bragg recommendations are followed. Specifically, it seems that Fast is reporting Siegbahn units (kX units), and that Duwez is using the absolute angstrom (10^{-8} cm), although even this assumption places Duwez's determination for c slightly higher than parameters reported by Fast,[1] Litton,[5] and by this laboratory.[3] Assuming a Vegard relation between the lattice parameters of Zr and Hf, and using Fast's data for these elements, the Duwez value would seem to fit the line at approximately 94 atomic percent hafnium rather than the low

TABLE I. Lattice parameters of some zirconium-hafnium alloys at room temperature. (In this table the Bragg (reference 6) recommended conversion relation, $1A = 10^{-8}$ cm $= kX/1.00202$ is used. The italicized values are the published ones and the symbols used where noted with these values are those that appear in the reference.)

Percent hafnium wt	Percent hafnium at	Temp °C	c A	c kX	a A	a kX	Reference
100	100	room	5.051	*5.041*A	3.193	*3.187*A	Fast[a]
100	100	room	*5.0569*A	5.0467	*3.1952*A	3.1888	Duwez[b]
100	100	24	5.0510	5.0409	3.1946	3.1881	This paper[c]
99.0	98.5	room	5.0524	*5.0422*A	3.1947	*3.1883*A	Litton[d]
98	96.4	26	5.0602	5.0500	3.1992	3.1928	This paper[c]
76	61.8	25	5.0944	5.0842	3.2108	3.2043	This paper[c]
2.4	1.2	25	5.1457	5.1353	3.2309	3.2244	Russell[e]
0	0	room	5.144	*5.134*A	3.231	*3.225*A	Fast[a]
0.01	0.005	25	5.1477	5.1374	3.2312	3.2247	Russell, unpublished

[a] See reference 1.
[b] See reference 2.
[c] See reference 3.
[d] See reference 5.
[e] See reference 4.

purity of 81 atomic percent hafnium that Fast assigns. Actually there seems to be a positive Vegard deviation for the high hafnium alloys, so that small zirconium additions can cause a rather large increase in lattice parameters.

* This work was performed at the M.I.T. Metallurgical Project under Contract No. AT(30-1)-981 with the U. S. Atomic Energy Commission.
[1] J. D. Fast, J. Appl. Phys. **23**, 350 (1952).
[2] P. Duwez, J. Appl. Phys. **22**, 1174 (1951).
[3] In a letter to Dr. A. R. Kaufmann of this laboratory from Dr. A. Nederlof of N. V. Philips Company, Eindhoven, Holland, June 27, 1949, it was stated that the hafnium content was calculated from the specific gravities of these alloys, since experiments have shown that volumes were additive, and are as follows:

Alloy	Spec Wt
100 percent wt. Hf	13.28
98 percent wt. Hf	13.05
76 percent wt. Hf	10.64

[4] R. B. Russell, "The Coefficients of Thermal Expansion for Zirconium," MIT Metallurgical Project Report No. MIT-1073, under Atomic Energy Commission Contract No. AT-(30-1)-981 (October 19, 1951).
[5] F. B. Litton, J. Electrochem. Soc. **98**, 488 (1951).
[6] W. L. Bragg, J. Sci. Instr. **24**, 27 (1947).

New Dispenser Type Thermionic Cathode

ROBERTO LEVI
Philips Laboratories, Inc., Irvington-on-Hudson, New York
(Received November 24, 1952)

IN recent years there has been developed in Holland a dispenser type cathode with useful properties. This cathode, which has been termed the "*L* cathode," has been described in some detail by Lemmens, Jansen, and Loosjes,[1] and an example of a planar cathode of this type is shown in Fig. 1 (a). A development program under aken in these laboratories aimed primarily at methods for improving fabrication techniques has finally resulted in a variant of the *L* cathode (see Fig. 1 (b) for an example), which is being called the Philips "impregnated cathode." Interest in this cathode has developed to such a degree that it appears desirable to publish details about its structure and some of its inherent advantages.

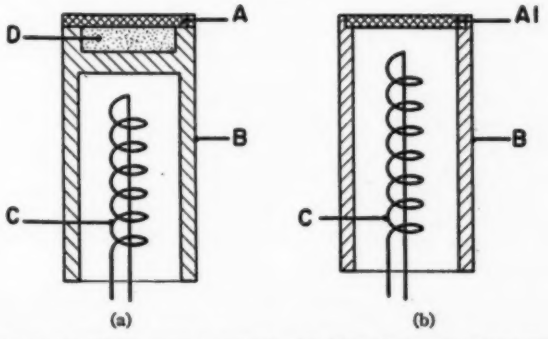


FIG. 1. (a) Planar type *L* cathode. (b) New planar type impregnated cathode. *A* = porous tungsten body, *A*1 = impregnated tungsten body, *B* = molybdenum body, *C* = heater, and *D* = cavity containing $(Ba-Sr)CO_3$.

The new cathode has eliminated the need for a large cavity by dispersing the alkaline earth material within the pores of the tungsten body. The most important result of the new structure is an extreme simplification in cathode construction and a vast extension of the possible sizes and shapes that can be fabricated. One of the reasons for this is that a technique also developed by the author, which permits the machining of tungsten to within the same tolerances as may be achieved in the machining of brass, is especially well suited to the fabrication of the new cathode. As examples of what may be achieved with the new construction, planar type cathodes have been made in sizes ranging from 0.010 inch to 0.750 inch in diameter; cylindrical cathodes have been made as small as 0.060 inch long, 0.030 inch i.d., 0.040 inch o.d., and as large as $1\frac{1}{4}$ inches long, 0.440 inch i.d., 0.500 inch o.d.; and a spherical emitting surface 0.155 inch in diameter with a 0.155-inch radius of curvature has been prepared.

Other important advantages of the new construction deriving from the elimination of the large cavity relate to thermal problems; these include more homogeneous temperature distribution across the emitting area; improved thermal efficiency; and more space for the heater, permitting the use of larger heaters and resulting in longer heater life, especially for small cylindrical cathodes.

The alkaline earth material dispersed throughout the tungsten plug in the new cathode is a mixture of normal and basic barium aluminates in contrast to the mixture of alkaline earth carbonates employed in the *L* cathode. The use of the aluminates yields the important advantage in processing that practically no gas is evolved so that the time required on the pumping system is extremely short. In addition, the usual advantages of the *L* cathode over alkaline earth oxide emitters are retained in the impregnated cathode.

Further details regarding this cathode will appear at a future date in publications from this laboratory.

The author is pleased to express his indebtedness to the Chemical Section for preparation of several impregnant materials, and to the X-Ray Section for their analysis. Technical assistance on the part of Mr. A. DePaul, Mr. J. Lambertson, Mr. G. Low, and Mr. M. Miller has likewise proved helpful.

[1] Lemmens, Jansen, and Loosjes, Philips Tech. Rev. **11**, 341 (1950).

A Ball of Fire Discharge with a Pronounced Current Saturation

G. K. MEDICUS
Components and Systems Laboratory, Wright Air Development Center, Wright-Patterson Air Force Base, Ohio
(Received October 2, 1952)

IN a diode structure, having a hot oxide hollow cathode and a large anode cap sealed to the glass envelope, a marked current saturation in a range of about 6-v anode drop was obtained, at a pressure of 250 μXe.

In Fig. 1 the geometry of the diode is shown; Fig. 2 gives an example of a saturation curve. With the anode temperature kept constant, the changes of the saturation current stayed below one percent in a voltage range of about 5 v.

The mechanism of this discharge, as was found out by probe measurements, is as follows: In the ball of fire which is located near the opening of the hollow cathode a steep potential peak exists. The electrons, which are drawn into this region from the (nonsaturated) hollow cathode, produce a very sharp peak of high plasma density there. The retarding field which surrounds the ball of fire on all sides keeps the electrons from directly reaching the

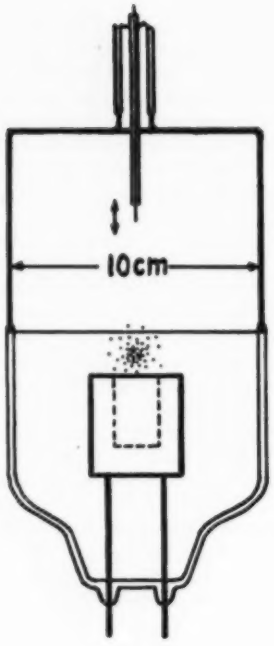


FIG. 1.

anode. Near the anode a fairly uniform plasma of low density is created. The electrons carrying the anode current reach the anode by diffusion against a retarding field. (Up to this point this mechanism in principle is well known.[1]) In the region of current saturation the anode behaves like a large plane probe which draws the electron saturation current from the adjacent plasma through a Langmuir sheath. The additional voltage drop in the region of current saturation occurs across this sheath, without influencing the plasma behind it at all, since the ionization in this thin sheath is negligibly small. At the upper knee of Fig. 2, however, ionization in this sheath sets in, so that the current rises above saturation. At a current of 0.82 amp (in this particular case) an abrupt breakdown occurs which changes the discharge fundamentally.

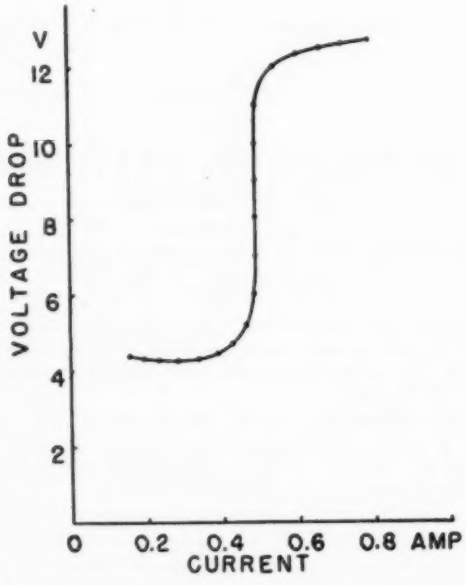


FIG. 2.

The saturation could not—or at least not clearly—be observed, when the minimum voltage drop of the diode was higher than about 6 v. Although the reasons for this unexpected behavior are not yet thoroughly understood, it is certain that "patchiness" of the anode plays an important part in this connection. Patchiness of the anode means spots of low work function which, for instance, form on it during processing of the cathode by evaporation of low work function material from the latter. These low work function spots on the anode, with increasing plate voltage, reach saturation current—and breakdown voltage—long before the rest of the anode. Consequently, on these spots ionization may set in long before the upper knee of Fig. 2 is reached. The positive ions flowing from these spots into the plasma increase the plasma density near them, causing a concentration of the anode current on them. This current concentration means further increase of ionization, so that an instability results which causes precocious breakdown of the Langmuir sheath near these spots. This mechanism is supported by the appearance of luminous channels originating at the anode in the case of poor saturation and otherwise similar conditions.

In order to make the saturation phenomenon clearly observable, it was found necessary to reduce the minimum voltage drop of the diode, for instance, by evaporating Ba from a getter wire onto the anode (and the cathode). By this means the minimum

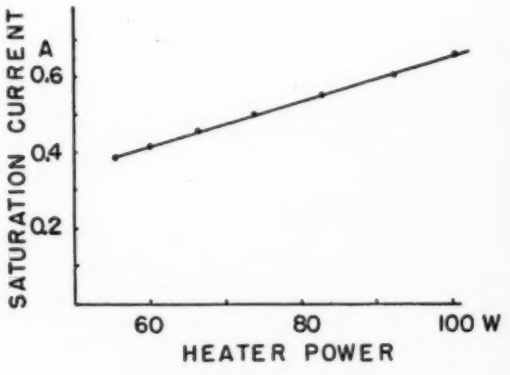


FIG. 3.

voltage drop can be lowered from initial values in the order of 6 v down to about 2 v. To what degree this lowering of the voltage drop is caused only by reducing and homogenizing the work function of the anode cannot be decided yet. Therefore, it cannot readily be explained yet why this saturation phenomenon in a ball of fire diode does not seem to have been observed before.

It should be noted that the transition from the nonsaturated to the saturated state causes no fundamental change of the potential distribution within the plasma, in spite of the fundamental difference in positive ion economy between these two states, which is to be expected on account of the absence or presence of a potential barrier for the ions near the anode. The comparatively large (nonemitting) outside area of the hollow cathode which constitutes a large sink for the ions, probably is responsible for this fact.

The saturation current—in the region where it exists—is a fairly linear function of the heating power of the cathode, as is shown in Fig. 3. This seems to indicate that the gas density near the ball of fire determines the plasma density peak in it and thereby the plasma density near the anode which yields the saturation current.

It is hoped that a more detailed paper on this subject can be published in the near future.

[1] See Malter, Johnson, and Webster, RCA Rev. 12, 415 (1951), where preceding literature is given.